THE UNIVERSITY *of* EDINBURGH

Host-adaptive evolution of

Staphylococcus aureus

Bethan Lowder

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2011

The research presented in this thesis is entirely my own work, except where otherwise stated. No part of this thesis has been submitted in any other application for a degree or professional qualification.

Bethan Lowder August 2010

Acknowledgements

I am grateful to Dr. Ross Fitzgerald for his support and guidance throughout the project, including numerous fruitful discussions both in and out of the lab. I would also like to thank the members of the Laboratory for Bacterial Evolution and Pathogenesis (the LBEP girls) Dr. Catriona Guinane, Gillian Wilson, Jeanette Bannoehr, Dr. Nouri Ben Zakour, Robyn Cartwright and Dr. Shruti Daga for being great friends and colleagues. In particular, thank you to Caitriona, whose assistance and inspiration have undoubtedly made this thesis what it is, to Nouri for sharing her wealth of bioinformatic expertise, and to Gill for being an excellent PhD buddy from day one. Members of the ZAP, MPRL and CF research groups have also been a pleasure to work with. My thanks, especially, to Allen Flockhart for his infectious enthusiasm and endless capacity to listen to me talk about my project.

Many thanks to Jon Bollback for his generous contribution to the selection study, and for many discussions and patient explanations. I am also grateful to The Gene Pool (The University of Edinburgh) for technical and bioinformatic assistance. In addition, isolates for this study were kindly donated for this analysis by J. Rodgers, G. Foster, D. Vancraeynest, W. Grubb, D. Bertolatti, S. Takeuchi, L. Guardabassi, J. Lindsay, J. Penades, T. Tollersrud, H. de Lencastre, J. Maurer, and M. Maier.

Last, but by no means least, I would like to thank my family for their unwavering moral support and interest in my work. I could not have done it without you. Thank you to Michael, Mum, Dad, Amy, Oliver, Gill, Granny, Christine, Peter, Jennifer and Andy.

Glossary

α	Alpha
β	Beta
BCO	Bacterial chrondonecrosis with osteomyelitis
BER	BLAST extend repraze
BHI	Brain and heart infusion
BLAST	Basic local alignment search tool
bp	Base pair
BSE	Bovine spongiform encephalopathy
CC	Clonal complex
CDS	Coding sequence
DNA	Deoxyribose nucleic acid
dH_2O	Distilled water
dN	Non-synonymous mutation
dS	Synonymous mutation
ECL	Enhanced chemiluminescent
g	Gram
h	Hour
HGT	Horizontal gene transfer
IEC	Immune evasion cluster
Ig	Immunoglobulin
IS	Insertion sequence
JCVI	J. Craig Venter Institute
kbp	Kilobase pair
kDa	Kilodalton
MEGA 4	Molecular evolutionary genetics analysis 4
MGE	Mobile genetic element
mg	Milligram
min	Minute
ml	Millilitre
MLEE	Multi locus enzyme electrophoresis

MLST	Multi locus sequence typing
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSCRAMM	Microbial surface components recognising adhesive matrix molecules
NCBI	National centre for biotechnology information
ω	Relative rate of non-synonymous to synonymous substitutions
OD_{600}	Opitical density at 600nm
ORF	Open reading frame
PAML	Phylogenetic analysis by maxiumum likelihood
PAUP*	Phylogenetic analysis using parsimony
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFGE	Pulsed field gel electrophoresis
RAST	Rapid annotation using subsystem technology
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
SaPI	Staphylococcal pathogenicity island
SARS	Severe acute respiratory syndrome
SCC	Staphylococcal cassette chromosome
Sdr	Serine dipeptide (Ser-Asp) repeat
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEED	An annotation database that integrates multiple genome sequences
SNP	Single nucleotide polymorphism
ST	Sequence type
TCS	Two component system
TLR	Toll-like receptor
TSA	Tryptic soy agar
TSB	Tryptic soy broth
TSS	Toxic shock syndrome
VISA	Vancomycin intermediate *Staphylococcus aureus*

VRSA	Vancomycin-resistant *Staphylococcus aureus*
µl	Microlitre
V	Volt
w/v	Weight in volume
v/v	Volume in volume

Contents

Abstract

Staphylococcus aureus is a notorious human pathogen associated with severe nosocomial and community-acquired infections. In addition, *S. aureus* is a major cause of animal diseases including skeletal infections of poultry and bovine and ovine mastitis, which are a large economic burden on the broiler chicken and dairy farming industries.

The population structure of *S. aureus* associated with humans has been well studied. However, despite the prevalence of *S. aureus* infections in broiler flocks, our understanding of the diversity of poultry *S. aureus* is very limited. In this study, multi-locus sequence typing was performed on 48 strains of *S. aureus* isolated from broiler chickens on farms in 6 countries on 4 different continents, in addition to 9 isolates from different species of reared game and wild birds in Scotland. This was followed by fine scale population genetic analysis of a subset of strains by single nucleotide polymorphism discovery. These studies reveal that the majority of *S. aureus* isolates from broiler chickens are the descendants of a single human-to-poultry host jump by a subtype of the worldwide human clonal complex 5 (CC5) clonal lineage unique to Poland. In contrast to human subtypes of the CC5 radiation, which demonstrate strong geographic clustering, the poultry CC5 clade was distributed in different continents, consistent with wide dissemination via the global poultry industry distribution network.

In order to establish the molecular basis for avian specificity in the CC5 poultry clade, whole genome sequences were determined for a sequence type 5 (ST5) poultry isolate from Ireland and a basal human associated ST5 MRSA strain from Poland. Sequence analysis revealed that the poultry CC5 clade has undergone genetic diversification from its human progenitor strain by acquisition of novel mobile genetic elements from an avian-specific accessory gene pool, and by the inactivation of several proteins important for human disease pathogenesis.

In order to examine the importance of positive selection in the adaptation of *S. aureus* to poultry and for *S. aureus* evolution, in general, genome-wide analysis of the ratio of synonymous to non-synonymous substitutions was performed on 30 strains from

humans and other animals, from diverse lineages. Positive selection has affected proteins from the majority of functional categories, resulting in diversification of the proteome, metabolome and replication capacity, which may be associated with adaptation of *S. aureus* to diverse environments. For several proteins, an elevated rate of non-synonymous substitutions unique to animal-associated lineages is consistent with a role for these proteins in host adaptation.

Taken together, the results of this study have determined the evolutionary history of a major new animal pathogen that has undergone rapid avian host adaptation and intercontinental dissemination. The data highlight the importance of gene acquisition and loss and positive selection in the adaptive evolution of *S. aureus*.

1. Introduction

1.1. *Staphylococcus aureus* clinical significance

Staphylococcus aureus is a Gram positive bacterium that infects humans and other animals (Table 1). It is a commensal organism that colonises the skin and anterior nares of approximately 30 % of the population (Gorwitz *et al.*, 2008), forming part of the resident microflora of many people. It is also, however, an opportunistic pathogen that invades tissues and causes disease if the skin or mucous membrane are breached, particularly where host immunity is suppressed (Davidson & Boucher, 2009). In humans, *S. aureus* is the most common cause of hospital-acquired infection and is estimated to result in clinical disease in approximately 2 % of all patient admissions (Emmerson *et al.*, 1996, Jones, 2003). Furthermore, the past two decades have seen the emergence of antibiotic-resistant strains in the community, which cause aggressive infections in young, otherwise healthy people (Okuma *et al.*, 2002, Vandenesch *et al.*, 2003, for review see Zetola *et al.*, 2005).

Invasion of a human host can cause a range of clinical conditions, from minor skin and soft tissue infections (Dryden, 2009) to life-threatening bacteraemia (Corey, 2009), infective endocarditis (Fowler *et al.*, 2005) and toxic shock syndrome (Fitzgerald *et al.*, 2001b, Shands *et al.*, 1980, Todd *et al.*, 1978). However, *S. aureus* infection is not restricted to humans, and bacteria can invade commercially important animal species including ruminants, poultry, rabbits, horses and pigs (Table 1).

Clinical presentation in animal species also varies, ranging from superficial impetigo in cattle (George *et al.*, 2008) and dermatitis in goats (Matthews, 1999) to severe abscesses in major organs, clinical mastitis and pyaemia (Table 1). Several conditions cause significant morbidity and mortality, resulting in substantial economic losses for the farming industry. For example, bovine mastitis reduces milk volume and quality costing approximately £200 million annually (UK), with *S. aureus* identified as one of the most

Table 1. Clinical presentation of *S. aureus* infection in different host species.

Host species	Diseases associated with *S. aureus* infection	References
Human	Bloodstream infections	Corey, 2009 Seybold *et al.*, 2006
	Skin and soft tissue infection	King *et al.*, 2006
	Toxic shock syndrome	Fitzgerald *et al.*, 2001b Musser *et al.*, 1990
Cattle	Mastitis	Biggs, 2009
	Udder impetigo	George *et al.*, 2008
Sheep	Mastitis	Watkins & Jones, 2007
	Tick pyaemia	Quinn *et al.*, 2002
	Benign folliculitis, impetigo, eczema	McNeil, 2007
Goat	Mastitis	Matthews, 1999
	Dermatitis, pustular dermatitis	Matthews, 1999
Poultry	Septic arthritis	Smyth & McNamee, 2008
	Subdermal abscesses (bumble foot)	Smyth & McNamee, 2008
	Bacterial chondronecrosis with osteomyelitis (BCO)	McNamee & Smyth, 2000

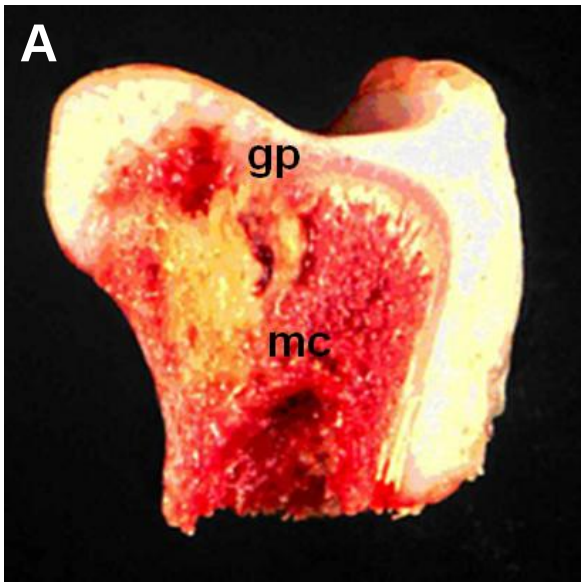
Host species	Diseases associated with *S. aureus* infection	References
Rabbit	Pododermatitis, subcutaneous abscesses and mastitis	Hermans *et al.*, 2003
	Septicaemia leading to abscess formation in lungs, liver and uterus	Hermans *et al.*, 2003
Swine	Botromycosis and impetigo of mammary glands	Quinn *et al.*, 2002 Taylor, 1999 van Belkum *et al.*, 2008
Horses	Wound/bite and postoperative infections	Leonard & Markey, 2008
Companion animal	Abscesses and infection of skin, eyes, ears, respiratory and genitourinary tracts and skeleton.	Cox, 2006 Leonard & Markey, 2008

frequent causative organisms (Biggs, 2009). In sheep and goats tick pyaemia is a major concern with up to 30 % of lambs affected each year (UK and Ireland), and infection can lead to septicaemia or abscess formation resulting in death (Quinn *et al.*, 2002). In regions where rabbit farming is widespread, a hyper-virulent clone of *S. aureus* has been identified. Clinical presentation is generally superficial podomermatitis and subcutaneous abscesses, though these can lead to secondary septicaemia and abscess formation in the lungs, liver and uterus (Hermans *et al.*, 2003). Infertility and morbidity rates reach 65 % to 80 % in affected flocks, with mortality up to 35 % (Hermans *et al.*, 2003).

1.1.1. Infection of poultry

Infectious diseases of chicken flocks are a major economic burden on the poultry industry. In commercial broiler chickens, *S. aureus* infection can cause septic arthritis, septicaemia, subdermal abscesses (bumble foot) and gangrenous dermatitis (Smyth & McNamee, 2008; Fig. 1). Notably, in the 1970s a new form of *S. aureus* infection of broiler poultry known as bacterial chondronecrosis with osteomylitis (BCO) was described (Nairn & Watson, 1972). Infection causes lesions of the femur and tibiotarsus which render the bird lame, limiting its ability to access food and water and resulting in weight loss and death within a few days (McNamee & Smyth, 2000).

Since it was first described, BCO has increased in frequency to become a leading cause of lameness in the broiler chicken industry (McNamee & Smyth, 2000, Nairn & Watson, 1972). In the 1960s and 70s lameness was predominantly associated with bone growth abnormalities, with arthritis and osteomyelitis appearing low down on a prioritised list of leg disorders (Ministry of Agriculture, Fisheries and Food, 1981). By the 1990s this situation had changed and the majority of cases were caused by *S. aureus* infection (McNamee *et al.*, 1998). The reasons for the emergence and subsequent increase in incidence of BCO among chickens are not known.



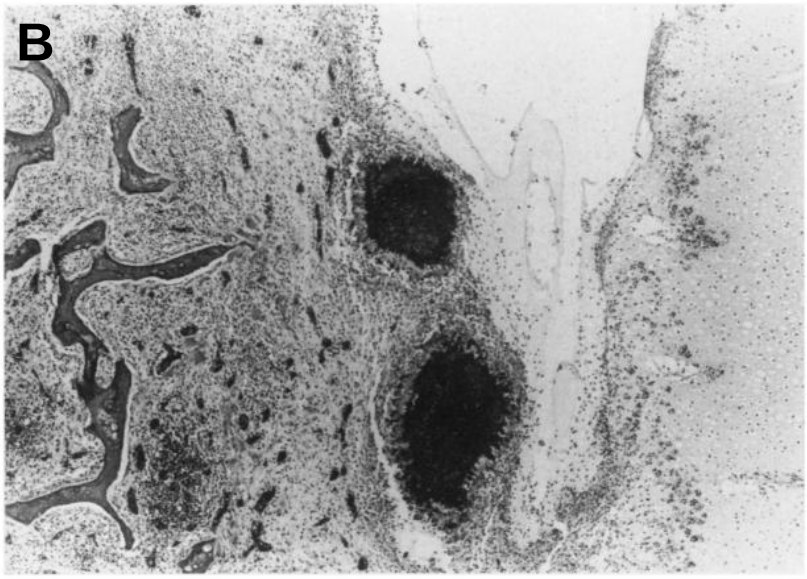


Figure 1. ***Staphylococcus aureus*** **infection of broiler chickens. (A)** Proximal end of the femur of a 28-day-old lame commercial broiler. Bacterial infection constitutes region of yellow tissue extending from growth plate region (gp) to medullary cavity (mc) (reproduced with permission from McNamee & Smyth, 2000; http://www.informaworld.com). **(B)** Chronic granulomatous lesions of osteomyelitis in the proximal tibiotarsus of an infected chicken. Bacteria are densely packed in the centre of lesions, alongside degenerated inflammatory cells and tissue debris (reproduced with permission from Mutalib *et al.*, 1983).

1.1.2. Antibiotic resistance

Antibiotic resistance has become a major obstacle to effective treatment of *S. aureus* infection. The first methicillin-resistant *S. aureus* (MRSA) strains were identified in the 1960s and are now prevalent across the US, Europe and Japan. In the UK, the proportion of MRSA isolates causing *S. aureus* bacteraemia peaked at approximately 45 % between 2001 and 2005, and elsewhere in Western and Southern Europe rates frequently exceed 25 % (Woodford & Livermore, 2009). Prevalence is higher in the US, with MRSA constituting approximately 55 % of laboratory *S. aureus* isolates examined in 2008 (Farrell *et al.*, 2009). The highest rates were recorded in Japan in the 1990s with approximately 60 % of strains displaying resistance (Goto *et al.*, 2009). Furthermore, methicillin-resistant strains have been isolated from cats, dogs, pigs, horses, cows and poultry (for review see Leonard & Markey, 2008), demonstrating that resistance is not restricted to human isolates.

The key mechanism for methicillin resistance in MRSA strains is the expression of penicillin-binding protein (PBP) 2a encoded on the *mec* mobile genetic element. The PBPs are a family of proteins that catalyse cross-linking (transpeptidation) reactions in the peptidoglycan layer of the staphylococcal cell, and β-lactam antibiotics such as methicillin inhibit these proteins. Though PBP2a has the structural motifs associated with penicillin binding found in other PBPs, β-lactam antibiotics do not bind effectively (Stapleton & Taylor 2002). Therefore, it is able to perform transpeptidation reactions in the presence of methicillin concentrations that would inhibit other PBPs. In some *S. aureus* isolates PBP2a is expressed at a relatively low concentration, however mutations in regulatory genes can lead to higher expression if this provides a selective advantage (Stapleton & Taylor 2002).

The first occurence of a clinical *S. aureus* isolate with reduced susceptibility to the antibiotic vancomycin was reported in Japan in 1997 (Hiramatsu, 1997), with further cases identified soon after in the US (Hageman *et al.*, 2001, Khurshid *et al.*, 2000, Rotun

et al., 1999). In 2002 a strain with full resistance was identified (Chang *et al.*, 2003). Such strains are referred to as vancomycin intermediate *S. aureus* (VISA) and vancomycin-resistant *S. aureus* (VRSA), respectively (Cosgrove *et al.*, 2004). VRSA strains remain very rare and the proportion of strains with intermediate resistance is much lower than for methicillin (typically <10 %; for review see Howden *et al.*, 2010). However, there are some notable exceptions, such as a hospital in Turkey where the rate of heterogeneous VISA strains increased from 1.6 % in 1998 to 32 % in 2001 (Sancak *et al.*, 2005).

Genome sequencing has led to considerable progress in understanding the molecular basis for resistant phenotypes, as a result of the comparison of genomes of antibiotic susceptible and resistant strains. Regions of the accessory genome important for resistance have been identified, demonstrating the importance of plasticity in the *S. aureus* genome and its ability to evolve rapidly by horizontal acquisition of advantageous mobile genetic elements (MGEs) (Holden *et al.*, 2004, Kuroda *et al.*, 2001). Furthermore, point mutations in housekeeping genes have been associated with resistance to quinolones (Holden *et al.*, 2010, Tanaka *et al.*, 2000). Specific nucleotide changes associated with resistance have also been pinpointed following scrutiny of sequential isolates taken at different time points throughout an infection (Mwangi *et al.*, 2007), or strains with differing levels of resistance (Neoh *et al.*, 2008).

1.2. Virulence and pathogenesis of *S. aureus*

S. aureus produces an array of proteins important for survival and pathogenesis, many of which have been directly linked to disease symptoms in the host. These include secreted extracellular proteins and those attached to the cell surface, which are involved in the key stages of infection: attachment to host cells and extracellular matrices, evasion of the host immune system, and invasion of host cells and/or tissues.

1.2.1. Attachment

A prominent family of proteins that enable attachment of bacteria to extracellular matrix and plasma proteins, promoting colonisation and dissemination throughout the host, are the microbial surface components recognising adhesive matrix molecules or MSCRAMMS (Foster & Hook, 1998, Clarke & Foster, 2006). The majority of MSCRAMMs are covalently attached to the cell wall peptidoglycan via the LPXTG motif at the C-terminus, adjacent hydrophobic residues and positively charged tail (Schneewind *et al.*, 1993). Comparative analysis of six *S. aureus* genomes suggests that this family contains 21 proteins (Roche *et al.*, 2003), which between them facilitate adhesion to a wide range of ligands (for review see Clarke & Foster, 2006).

For example, Staphylococcal protein A (SpA) is conventionally described as an immunoglobulin (Ig) G- and IgM-binding protein, but can also bind von Willebrand factor, a serum glycoprotein produced at sites of endothelial damage (Hartleib *et al.*, 2000), and TNFR1, a receptor for tumour necrosis factor A (Gomez *et al.*, 2004). Two fibronectin-binding proteins, FnbpA and FnbpB, have also been identified (Greene *et al.*, 1995), though FnbpA also binds fibrinogen (Wann *et al.*, 2000) and both proteins bind elastin (Roche *et al.*, 2004). Finally, the Sdr family of MSCRAMMS, which have a characteristic Ser-Asp dipeptide repeat region and a conserved structure, include fibrinogen-binding proteins ClfA and ClfB (McDevitt *et al.*, 1997, Ní Eidhin *et al.*, 1998).

1.2.2. Evasion of host defences

Protection from the host immune system is achieved with a range of proteins that enable bacteria to inhibit neutrophil chemotaxis, kill leukocytes and resist phagocytosis, and confer protection against antimicrobial peptides and lysozyme (Foster, 2005). Each of these contributes to the ability of *S. aureus* to invade many tissues and host species by overcoming innate and acquired immune responses. Well-studied examples include capsule polysaccharides and SpA, both of which inhibit phagocytosis (Gemmell *et al.*, 1991, O'Riordan & Lee, 2004). Extracellular enzymes are also produced, including

proteases, lipases and fatty acid modifying enzyme (FAME), which inactivate the compounds produced by the host with the intention of disrupting bacterial activity (Dubin, 2005, Mortensen *et al.*, 1992, Kapral *et al.*, 1992). Exotoxins produced include superantigens, such as toxic shock syndrome toxin 1 (TSST-1), which stimulate non-specific T-cell proliferation (Choi *et al.*, 1989, Mourad *et al.*, 1989), and cytolytic toxins which irreversibly damage the membranes of host cells (Foster, 2005). For example, α toxin forms β-barrel pores in the cytoplasmic membrane of the target cell, promoting lysis of neutrophils that would otherwise engulf and destroy bacteria (Montoya & Gouaux, 2003).

1.2.3. Invasion of host tissues

Many of the proteins implicated in evasion of host defences are also critical for bacterial invasion as they damage host tissues. For example, exfoliative toxins target a desmosomal glycoprotein responsible for maintaining cell-to-cell adhesion in the superficial epidermis, leading to the characteristic separation of the epidermal layers seen in scalded skin syndrome (Ladhani, 2003). Similarly, hyaluronate lyase degrades mucopolysaccharide hyaluronic acid, a major component of human and animal connective tissue (Makris *et al.*, 2004), and staphylokinase facilitates activation of plasminogen to plasmin, degrading fibrin clots and promoting host invasion (Collen, 1998).

1.2.4. Regulation of virulence factors

Few bacterial genes are expressed constitutively, particularly those of the accessory genome (Novick, 2006), and regulation of virulence factor production depends on a combination of environmental and genetic factors. In *S. aureus*, a number of two component systems (TCSs) have been identified, each consisting of a transmembrane sensor and a transcription factor known as a response regulator, which are usually encoded on the same operon (Projan & Novick, 1997). Using such systems, bacteria are able to respond to environmental cues such as pH, temperature, glucose concentration or

cell density with expression of particular virulence factors, ensuring that they are only produced at the appropriate stage of infection (Projan & Novick, 1997).

The best characterised TCS is the *agr* (accessory gene regulator) locus, which plays a role in the expression of numerous virulence factors (Fournier, 2008). Others include *sae* (*S. aureus* exoprotein expression), *arl* (autolysis related locus) and *srr* (staphylococcal respiratory response) (Novick, 2006). Though these TCS modules can affect target genes directly, they also interact with intermediary transcription factors (Novick, 2006). Several have been characterised, including *rot* (repressor of toxins), *sar (*staphylococcal accessory regulator*) A, sarR, sarS, sarT, sarU, sarV, tcaR* (teicoplanin-associated locus regulator) and *rat* (regulator of autolytic activity) (Novick, 2006).

1.3. Population genetics of *S. aureus*

Population genetic analysis can provide insight into the evolutionary history of a pathogen, shedding light on important aspects of molecular evolution, niche adaptation, population divergence and the relationship between host and microbe (Maiden & Urwin, 2006).

1.3.1. Phylogeny of the *S. aureus* species

Phylogenetic population structure across the *S. aureus* species has been examined in many studies, and typing techniques vary depending on the degree of differentiation required. Early work implemented biotyping systems, which group strains according to phenotypic characteristics (Devriese, 1984, Devriese & Oeding, 1976, Isigidi *et al.*, 1990). Further discrimination was achieved with multi locus enzyme electrophoresis (MLEE), which provided evidence that isolates from cattle or sheep with mastitis and those from humans with toxic shock syndrome originated from distinct clonal groups (Musser *et al.*, 1990). Pulsed field gel electrophoresis (PFGE) has also been used to

suggest clonal relatedness between different bacterial populations. For example, isolates from hospitals throughout Poland cluster into two clonal groups when analysed using PFGE (Leski *et al.*, 1998). However, such techniques are less effective at determining evolutionary relatedness, and data generated in different laboratories are not readily comparable.

As a successor to such techniques, single nucleotide polymorphism (SNP) analysis is highly discriminatory and now widely employed. The phylogenies of at least 25 species, including *Escherichia coli* (Wirth *et al.*, 2006), *Haemophilus influenzae* (Meats *et al.*, 2003) and *Streptococcus pneumoniae* (Zhou *et al.*, 2000), have been reconstructed using SNPs in small numbers of genes, using a technique called multi-locus sequence typing (MLST). Genes selected for MLST are housekeeping genes, thus nucleotide differences between two strains are thought to be the result of genetic drift and reflect evolutionary distance rather than differing selective pressures (Maiden *et al.*, 1998). Strains are assigned a sequence type (ST) number based upon the combination of alleles at seven loci, and closely related genotypes grouped into clusters known as clonal complexes (CCs). An MLST system has been developed for *S. aureus* (Enright *et al.* 2000; Feil *et al.* 2003) and numerous studies reveal extensive diversity, with more than 1400 STs identified to date.

Several investigations have provided evidence that the *S. aureus* species is highly clonal (Feil *et al.*, 2003, Lindsay *et al.*, 2006). Point mutation is the primary mechanism of evolution of the core genome and homologous recombination has not had a major impact (Feil *et al.*, 2003), in contrast to other bacterial species such as *Helicobacter pylori* and *Listeria monocytogenes* where recombination has played a pivotal role (Suerbaum *et al.*, 1998, Orsi *et al.*, 2008). Studies of large collections of *S. aureus* isolates have identified several distinct CCs with a high degree of separation between each (Feil *et al.* 2003; Lindsay *et al.* 2006). This has been attributed, at least in part, to the *Sau*1 type I restriction-modification system (Waldron & Lindsay, 2006), an important mechanism that limits the uptake of genetic information from an external

source. The system comprises a modification enzyme that chemically modifies native DNA and a restriction enzyme that digests un-modified DNA. The *Sau*1 genes are specific to each clonal group, and this is thought to moderate the frequency of recombination between lineages (Waldron & Lindsay, 2006).

Some clonal lineages are responsible for disease in multiple geographical locations worldwide, though local frequencies vary significantly and some countries have a high proportion of unique CCs. A major study of UK human isolates identified at least 11 lineages, including the major clones CC1, CC5, CC8, CC30 and CC45 (Feil *et al.*, 2003, Lindsay *et al.*, 2006; Fig. 2). Dutch and US isolates also belong to these lineages with CC30 and CC45 dominating (Melles *et al.*, 2008; Fig. 2), though community-acquired MRSA isolates in the USA frequently belong to CC8 (Tenover *et al.*, 2006; Fig. 2). Data from other continents are sparse, though several recent studies have begun to

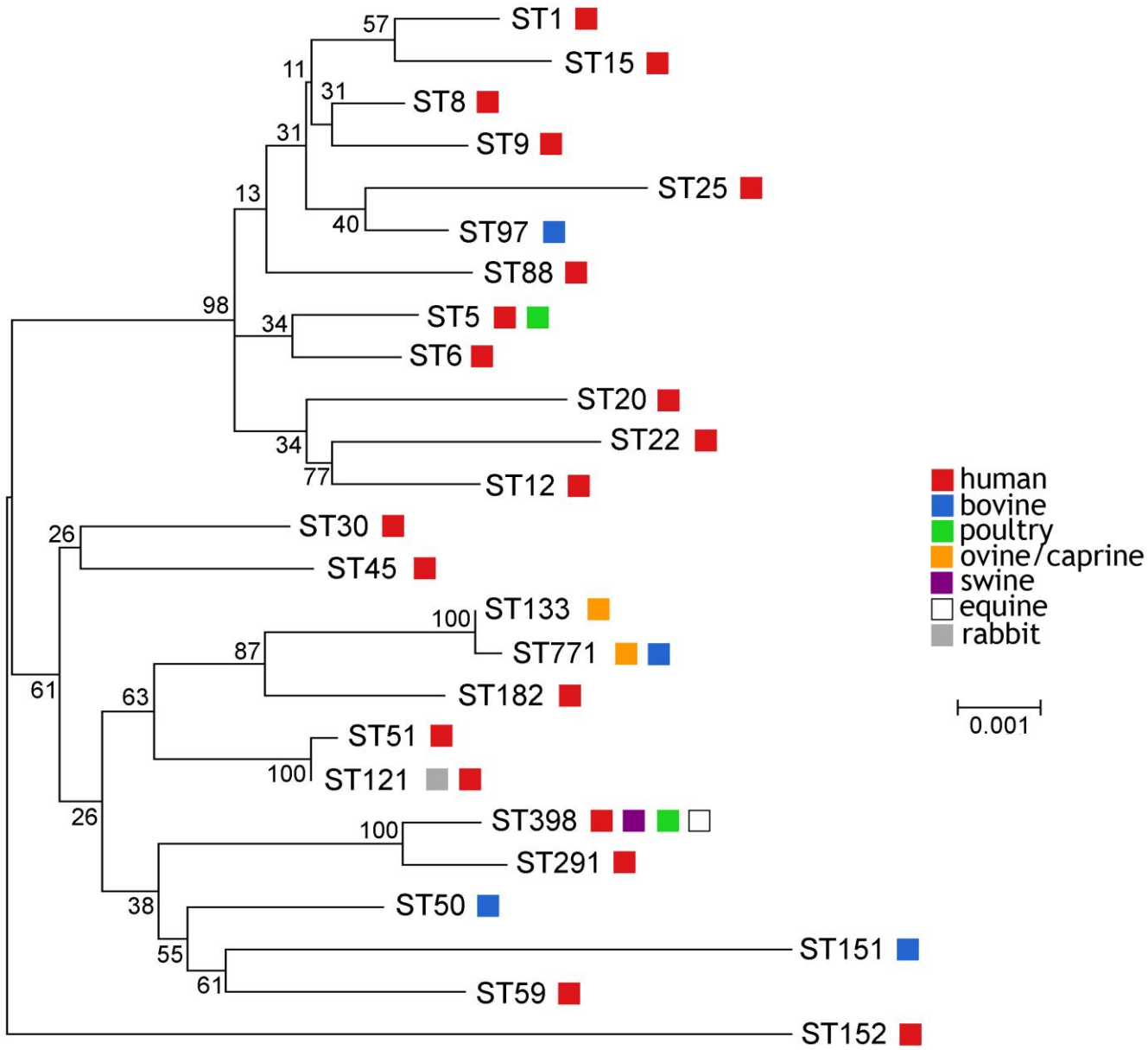

ST1
ST15
ST8
ST9
ST25
ST97
ST88
ST5
ST6
ST20
ST22
ST12
ST30
ST45
ST133
ST771
ST182
ST51
ST121
ST398
ST291
ST50
ST151
ST59
ST152
57
11
31
31
13
40
98
34
34
77
26
100
87
61
63
100
26
100
38
55
61
human
bovine
poultry
ovine/caprine
swine
equine
rabbit
0.001

Figure 2. The *S. aureus* species has diverged into multiple distinct evolutionary lineages which are predominantly host-restricted, with some notable exceptions. Phylogenetic tree constructed using concatenated sequences of seven housekeeping genes, as used in MLST system. STs included in analysis represent major lineages of human associated strains in several countries worldwide (Farrell *et al.*, 2009, Feil *et al.*, 2003, Lindsay & Holden, 2006, Melles *et al.*, 2008, Ruimy *et al.*, 2009, Ruimy *et al.*, 2008). Also included are major clades associated with cattle (Hasman *et al.*, 2010, Smith *et al.*, 2005, Smyth *et al.*, 2009, Sung *et al.*, 2008), small ruminants (Smyth *et al.*, 2009, Sung *et al.*, 2008), pigs (van Belkum *et al.*, 2008, Van Den Broek *et al.*, 2009), horses, rabbits and poultry (Smyth *et al.*, 2009). Coloured boxes indicate host association. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to branches.

address this issue. Strain collections from France, Algeria, Maldova and Cambodia all contain representatives of common worldwide clones but, in contrast to other locations, CC121 isolates constitute a large proportion of Cambodian strains and CC6 strains are commonly found in France, Algeria and Cambodia (Ruimy *et al.*, 2009; Fig. 2). Isolates from Mali also belong to the most prevalent lineages, though the frequency of ST152 PVL-positive community-acquired strains is also high (Ruimy *et al.*, 2008; Fig. 2). Finally, a Chinese strain collection differs from the global trend as it is dominated by strains of CC121, CC398, CC59 and CC182 (Fan *et al.*, 2009; Fig. 2).

Of note, MLST fails to distinguish between isolates of asymptomatic carriage and those of invasive disease (Feil *et al.*, 2003, Lindsay *et al.*, 2006), supporting the hypothesis that *S. aureus* is an opportunistic pathogen. The same strains that colonise the skin and respiratory tract can invade other tissues should the opportunity arise. This results in a paradoxical situation when considering bacterial adaptation and fitness of different strains with differing traits, as bacteria require genes for life as a commensal organism as well as those that enhance virulence in the disease state (Brown *et al.*, 2006).

Extending the principle of sequence-based typing to encompass many more genes, genome-wide SNP studies provide greater resolution within highly monomorphic taxa that display limited sequence diversity (Achtman, 2008). Thus, phylogenetic reconstruction is possible within a single CC.

Nubel *et al.* (2008) used such an approach to dissect CC5, a lineage responsible for hospital-acquired infection worldwide. They performed SNP discovery followed by typing of a global collection of 135 isolates, revealing several important features of the CC5 lineage. Firstly, subtypes within CC5 exhibit a high degree of geographical restriction suggesting that wide dispersal followed by fixation in a local population is a rare occurrence (Nubel *et al.*, 2008). Secondly, the staphylococcal cassette chromososme SCC*mec* responsible for antibiotic resistance has been imported into the CC5 group on at least 23 separate occasions (Nubel *et al.*, 2008), indicating that MRSA

is the result of frequent horizontal gene transfer (HGT) as opposed to expansion of a small number of resistant clones. Several other studies have reached similar conclusions about MRSA (Enright *et al.*, 2002, Fitzgerald *et al.*, 2001b, Musser & Kapur, 1992), contradicting the previous hypothesis that a resistant strain originated only once and spread worldwide (Kreiswirth *et al.*, 1993).

In order to further investigate one of the CC5 sub-lineages, Nubel *et al.* recently utilised the same approach to investigate American and European strains of ST225, a single locus variant of ST5 (Nubel *et al.*, 2010). Reconstruction of its evolutionary history revealed an ancestral population in the US with relatively high nucleotide diversity, but a more homogeneous population in Central Europe (Nubel *et al.*, 2010). This indicates a recent founder event from the US bacterial population to Germany and subsequent expansion within European hospitals, beginning around 1995 (1991 to 1999) (Nubel *et al.*, 2010).

SNP studies may also employ high-throughput sequencing of multiple genomes to highlight genome-wide single nucleotide variation, as with investigation of the ST239 lineage (Harris *et al.*, 2010). This study revealed that, as in the CC5 lineage, geographical restriction predominates but is interrupted by occasional intercontinental spread (Harris *et al.*, 2010). For example, ST239 strains caused two separate waves of healthcare-associated disease in Portugal, the first in the early 1990s and the second in 1997. Isolates from the second wave cluster with Brazilian isolates, suggesting that this outbreak resulted from the introduction of a South American strain to Portugal (Harris *et al.*, 2010). Though apparently rare, such founder events are a crucial component of strain emergence in a novel region.

Harris *et al.* highlighted the need for global surveillance strategies that include such a high degree of discrimination between isolates, in order to detect the introduction of clones to a new location. Furthermore, SNP typing studies such as those discussed here

can improve our understanding of the epidemiology of transmission, thereby increasing the effectiveness of intervention strategies in a particular setting (Harris *et al.*, 2010).

Importantly, population genetic analysis at any level may also tell us a great deal about the host. Though such a study has not been performed in *S. aureus*, interesting findings have been generated with analysis of *Helicobacter pylori*. As this species colonises the stomachs of half of the global human population and exhibits considerable genetic diversity, analysis of isolates can distinguish between closely related human groups more effectively than analysis of the human genome (Falush *et al.*, 2003, Wirth *et al.*, 2004). Studies revealed a population structure in Ladakh (India) consistent with known migration patterns (Wirth *et al.*, 2004), and structure in Africa, Asia and Europe that reflects events such as the neolithic introduction of farming to Europe and the slave trade (Falush *et al.*, 2003). Furthermore, comparison of *H. pylori* with its feline associated relative *H. acinonychis* provided evidence for a host jump from humans to felines within the last 200,000 years (Eppinger *et al.*, 2006).

1.3.2. Host specificity

Population genetic studies of *S. aureus* populations have tended to focus on human associated strains, and until recently few have included large collections of animal isolates. However, a phenotypic typing system describing distinct biotypes for strains from a human, poultry, bovine or ovine host has been in place for more than 25 years (Devriese, 1984, Isigidi *et al.*, 1990). In recent years, this approach has been superceded by molecular genetic approaches, revealing the existence of several host-adapted evolutionary lineages.

For example, *S. aureus* infection of the bovine udder leading to mastitis is a major economic burden on the dairy farming industry (Biggs, 2009), which has prompted research into the phylogenetic origin of bovine strains. As discussed above, strains isolated in a range of countries from milk or from the site of infection have been typed using MLEE, which suggested a separate clonal origin for human and bovine isolates

(Musser *et al.*, 1990). These also segregate by PFGE pulsotype (Zadoks *et al.*, 2000), though in a separate study skin isolates from both species belonged to the same group suggesting transmission during milking (Zadoks *et al.*, 2002). More recently, MLST has been performed, sometimes combined with other techniques such as comparative genomic hybridisation using microarray (Monecke *et al.*, 2007, Sung *et al.*, 2008) or *spA*, *sas* and *agr* typing (Hasman *et al.*, 2010, Smyth *et al.*, 2009). It has emerged that ST151, ST97, ST771 and CC50 are common bovine associated lineages (Hasman *et al.*, 2010, Smith *et al.*, 2005, Smyth *et al.*, 2009, Sung *et al.*, 2008; Fig. 2). Interestingly, ST151 demonstrates increased virulence and is hypersusceptible to the acquisition of vancomycin resistance genes from Enterococcus species (Guinane *et al.*, 2008, Sung *et al.*, 2008), indicating the emergence of a virulent subclone of *S. aureus* in cattle. In addition, genome sequencing of an ST151 isolate (strain RF122) and whole genome microarray analysis revealed a set of molecular genetic features unique to bovine-associated *S. aureus*, many of which are implicated in bovine mastitis pathogenesis (Herron-Olson *et al.*, 2007).

Mastitis is also a problem in other ruminant species, and analysis of isolates from ovine and caprine hosts reveals a similar pattern. The majority of isolates are from non human-associated lineages, in this case ST771 and ST133 (Smyth *et al.*, 2009, Sung *et al.*, 2008; Fig. 2). ST771 appears to be ruminant-specific as it is associated with cattle, sheep and goats (Sung *et al.*, 2008). ST133 is found in all ruminants but the vast majority of strains isolated from goats and sheep belong to this clonal group (Sung *et al.*, 2008, Guinane *et al.*, 2010).

Isolates from other animals, however, present a different scenario. The CC398 MRSA lineage is frequently isolated as a commensal of pigs but also seems to be readily transmitted to humans in the pig farming environment (van Belkum *et al.*, 2008, Van Den Broek *et al.*, 2009); Fig. 2). Typing of horse and poultry isolates suggests that CC398 strains may be transmitted to other livestock and companion animals (Persoons *et al.*, 2009, Van den Eede *et al.*, 2009, Walther *et al.*, 2009). Furthermore, a number of

horse, rabbit and chicken isolates have been found to be from lineages commonly associated with humans (CC8, CC121 and CC5, respectively) (Hasman *et al.*, 2010, Moodley *et al.*, 2006, Smyth *et al.*, 2009, Walther *et al.*, 2009). Smyth *et al.* postulated that this may indicate a jump from animal to human or vice versa. Transmission has also been documented between humans and companion animal species (Leonard & Markey, 2008, Moodley *et al.*, 2006, Strommenger *et al.*, 2006) but these may represent transient infections rather than the emergence of *S. aureus* strains adapted to dogs and cats.

It is clear that further investigation is required to determine the extent to which cross-species transmission of *S. aureus* strains occurs and whether this represents zoonoses or the emergence of strains which have adapted to a new host.

1.4. The *S. aureus* genome

Whole genome sequencing has revolutionised the field of microbiology. Analysis and comparison of complete genome sequences can reveal the molecular basis for phenotypic traits, and shed light on the processes driving bacterial evolution. Such studies complement and build upon the results of population genetic analysis, revealing molecular changes associated with adaptation within a particular lineage, or to a particular niche.

1.4.1. *S. aureus* genome structure

At the time of writing, the genome sequences of 30 strains of *S. aureus* have been published or released into the public domain and sequences are being completed at an ever increasing rate. This wealth of data has been exploited in a number of comparative studies. Sequenced *S. aureus* genomes range in size from 2,742,531 bp to 3,075,806 bp. The genome consists of one circular chromosome and up to 3 smaller circular plasmids. Approximately 2700 coding sequences are encoded depending on the strain, in addition to structural and regulatory RNAs, and the majority are annotated with a function though

many of these are putative based upon homology with known proteins or domains (Holden & Lindsay, 2008). Encoded proteins fall into one of multiple functional categories depending on their involvement with growth, replication and survival of cells and interaction with the host (Fig. 3). Cell envelope constituents comprise the greatest portion (20.3 % of total coding sequences), with transport/binding proteins (9.2 %) and MGE-associated genes (7.2 %) also prevalent.

Genome structure is largely conserved and consists of a core genome, which displays high nucleotide identity between strains, interrupted by regions of difference of approximately 500 bp to 50 kbp in size (Fig. 4). Bacterial genome sequencing projects traditionally divide the genome into core and accessory portions, but for *S. aureus*, Lindsay *et al.* (2006) further differentiated the accessory genome into MGE and a core-variable region. Core-variable genes display greater variation than core genes, but are typically stable and transferred vertically within particular clonal lineages (Lindsay *et al.*, 2006).

1.4.1.1. The core genome

Comparison of multiple *S. aureus* strains from different lineages by comparative genomic hybridiation studies revealed a core genome comprising approximately 77 % of the total number of open reading frames (ORFs) (Fitzgerald *et al.*, 2001b, Lindsay *et al.*, 2006). The core genome is highly conserved, with >95 % homology between orthologous pairs of genes in different strains (Lindsay *et al.*, 2006). Sequence divergence in core genes is due largely to point mutation (Feil *et al.*, 2003), and examples of gradual adaptive change have been identified following comparison of sequential clinical isolates (Mwangi *et al.*, 2007). A total of 35 SNPs associated with evolution of antibiotic resistance were identified in *S. aureus* JH1 to JH9 strains, suggesting that a limited number of mutations can lead to important phenotypic changes in this species (Mwangi *et al.*, 2007). Homologous recombination may also lead to diversification of the core genome, though studies suggest that recombination is >15 fold less likely (Feil *et al.*, 2003).

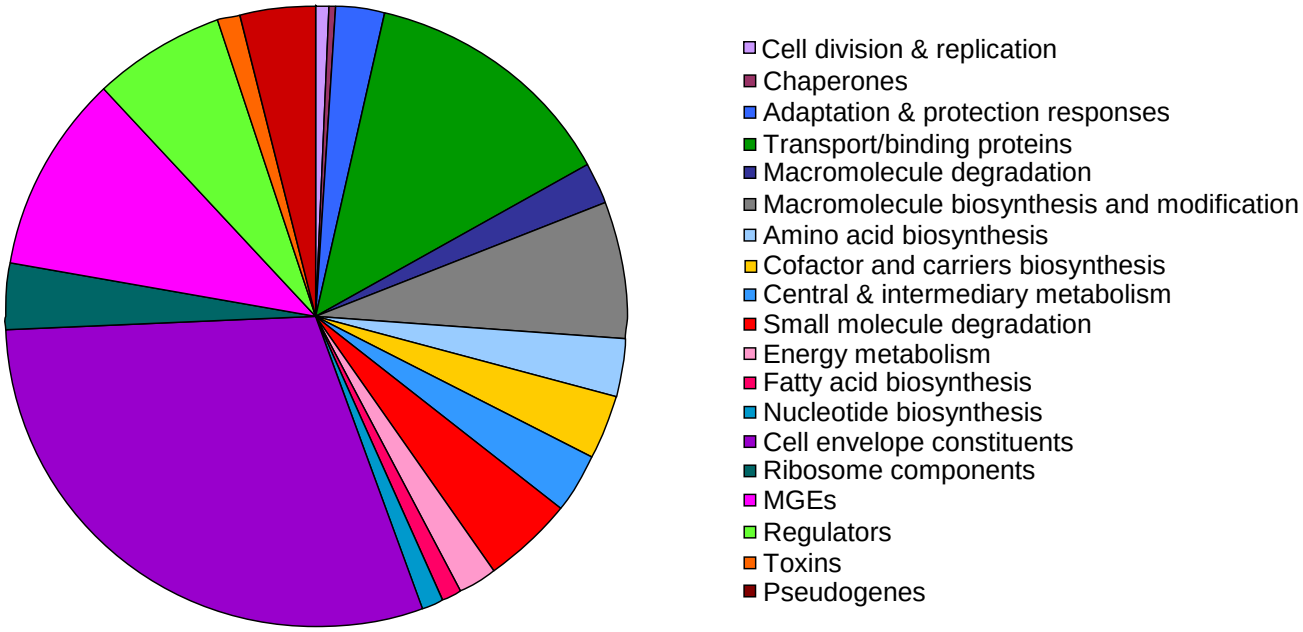


Figure 3. Coding sequences of the *S. aureus* genome encode proteins with roles in a variety of cellular processes. Genes encoding proteins involved in production of cell envelope, transport/binding proteins and MGE comprise greatest proportions of the genome, relative to other individual functional categories. Data from Holden & Lindsay (2008).

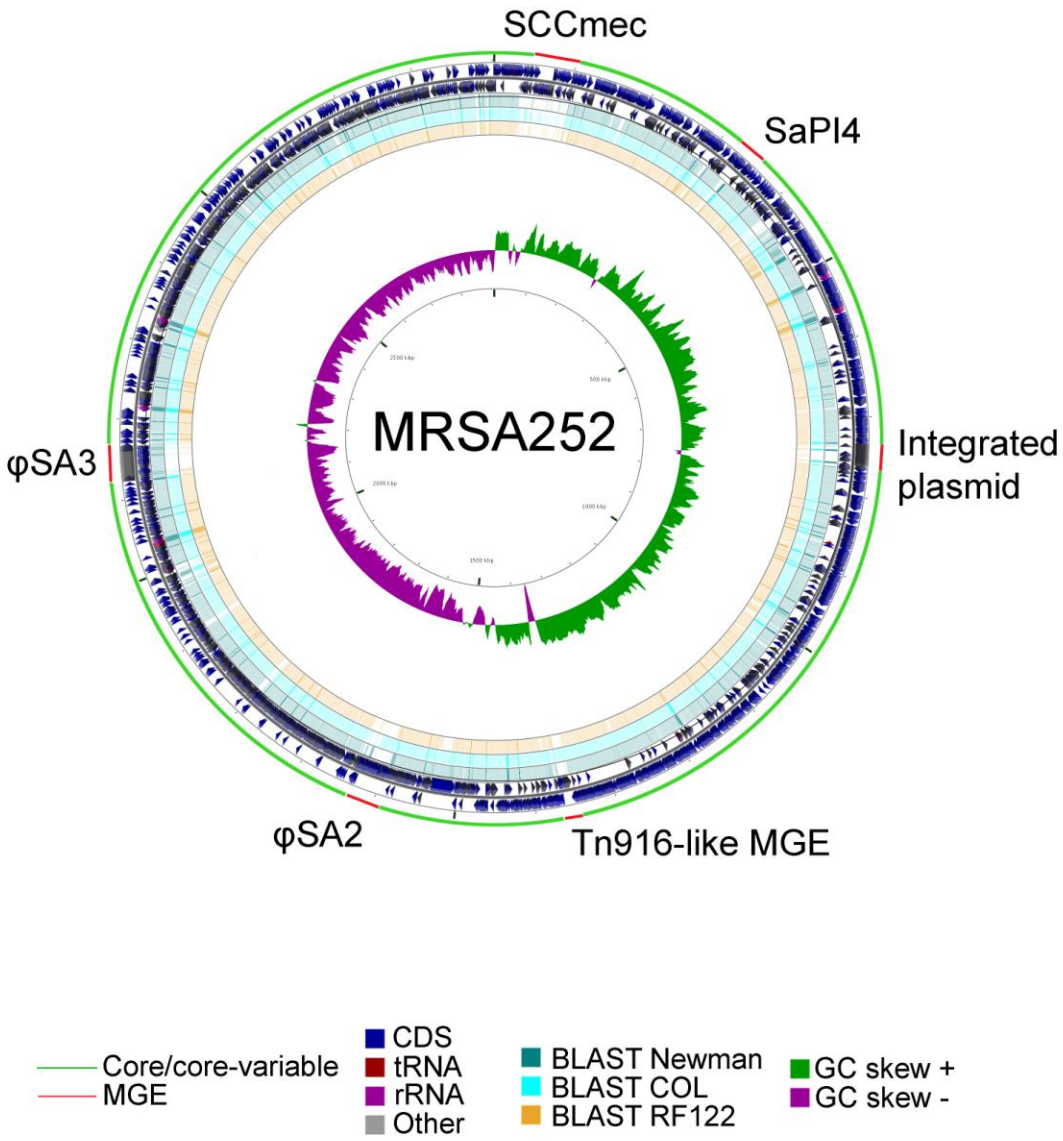


Figure 4. Structure of the *S. aureus* genome. Genome sequence of *S. aureus* strain MRSA252 with results of comparative nucleotide BLAST analysis against genomes of different strains. Outer circle (circle 1) indicates core/core-variable regions (green) interrupted by MGE (red). Circles 2 and 3 indicate CDS, structural and regulatory RNA and other genome features. Circles 4 to 6 display BLAST results, with coloured regions indicating high homology (e-value cutoff 0.1, identity >30 %) and white regions indicating no match. Circle 7 displays GC skew and inner circle displays size (kbp).

1.4.1.2. The core-variable genome

The core-variable region comprises approximately 10 % to 12 % of the *S. aureus* genome (Lindsay *et al.*, 2006). Some core-variable genes are present only in a subset of strains, commonly in accordance with CC (Lindsay *et al.*, 2006). Regions carrying from 1 to 9 genes, including many that encode imporant virulence determinants, may be inserted or deleted. Examples include the staphylococcal accessory regulator T gene *sarT* which regulates expression of virulence factors such as α-toxin (Schmidt *et al.*, 2001), and the staphylococcal anchored surface G gene *sasG* which promotes adherence to desquamated nasal epithelial cells and formation of biofilms (Corrigan *et al.*, 2007).

Alternatively, core-variable genes may be present in all strains but display substantial variation (Lindsay *et al.*, 2006). For example, FnbpA, an important virulence factor implicated in binding of host tissue components and foreign objects (Greene *et al.*, 1995) and in aggregation of platelets (Heilmann *et al.*, 2004), has a central region of approximately 145 bp which varies substantially between strains (Lindsay *et al.*, 2006). In addition, different variants of the *coa* gene encoding coagulase, a secreted protein that converts fibrinogen to fibrin (McDevitt *et al.*, 1992), have been identified in different CCs (Lindsay *et al.*, 2006).

1.4.1.3. Mobile genetic elements

MGE are an important component of the *S. aureus* genome and, as in many bacterial pathogens, often carry virulence or resistance determinants (Lindsay, 2008). Their distribution and the function of MGE-encoded proteins are widely studied due to their association with disease symptoms.

Bacteriophages

Transduction by bacteriophage is an important component of *S. aureus* evolution and some groups of prophage, such as the β-converting phage family (van Wamel *et al.*, 2006), have been well studied due to their role in conferring an advantageous phenotype to the host strain. Genome sequencing projects have revealed several prophage

sequences and many more have been identified independently (Iandolo *et al.*, 2002, Kwan *et al.*, 2005, Vybiral *et al.*, 2003; see also http://www.ncbi.nlm.nih.gov/genome). Relatedness of different phage types has been assessed by DNA hybridisation, restriction endonuclease profiling and virion protein profiling (Iandolo *et al.*, 2002, Lee & Stewart, 1985, Stewart *et al.*, 1985) as well as, more recently, whole phage sequencing (Bae et al 2006, Goerke et al 2009, Kwan et al 2005). Modern classification systems are commonly based on sequence variation across part of or the entire phage, and PCR typing systems have been developed based on sequence variation in the integrase and other genes (Goerke *et al.*, 2009, Pantucek *et al.*, 2004).

The majority of published *S. aureus* bacteriophages belong to the *Siphoviridae* family of temperate, tailed bacterial viruses (Goerke *et al.*, 2009), also referred to as class II phage (Kwan *et al.*, 2005). Phage genomes are generally around 40 kbp in length and exhibit a mosaic structure, though there is typically a conserved arrangement of functional modules for lysogeny, DNA replication, regulation of transcription, packaging and head, and lysis (Fig. 5). Other phage types identified in *S. aureus* include small (<20 kbp) class I phage as well as large (>125 kbp) class III phage which are members of the *Myoviridae* family (Kwan *et al.*, 2005). When a phage integrates into the chromosome it becomes a stable prophage and will replicate with the rest of the genome, being transferred vertically to daughter bacterial cells (McShan & Ferretti, 2007). The prophage can be induced by stress-inducing factors such as ultraviolet light and some antibiotics, which cause it to excise and replicate as an independent unit (McShan & Ferretti, 2007). Over time, the capacity to mobilise may be lost and all or part of the prophage sequence can become stably associated with the chromosome. Recombination and other mechanisms of genome degradation can lead to loss of functional structural genes whilst retaining those with a function important for the bacterium.

Staphylococcal pathogenicity islands (SaPIs)

Staphylococcal pathogenicity islands (SaPIs) are MGEs of around 15 kbp to 17 kbp, often encoding virulence genes. One or more SaPI has been found in almost all

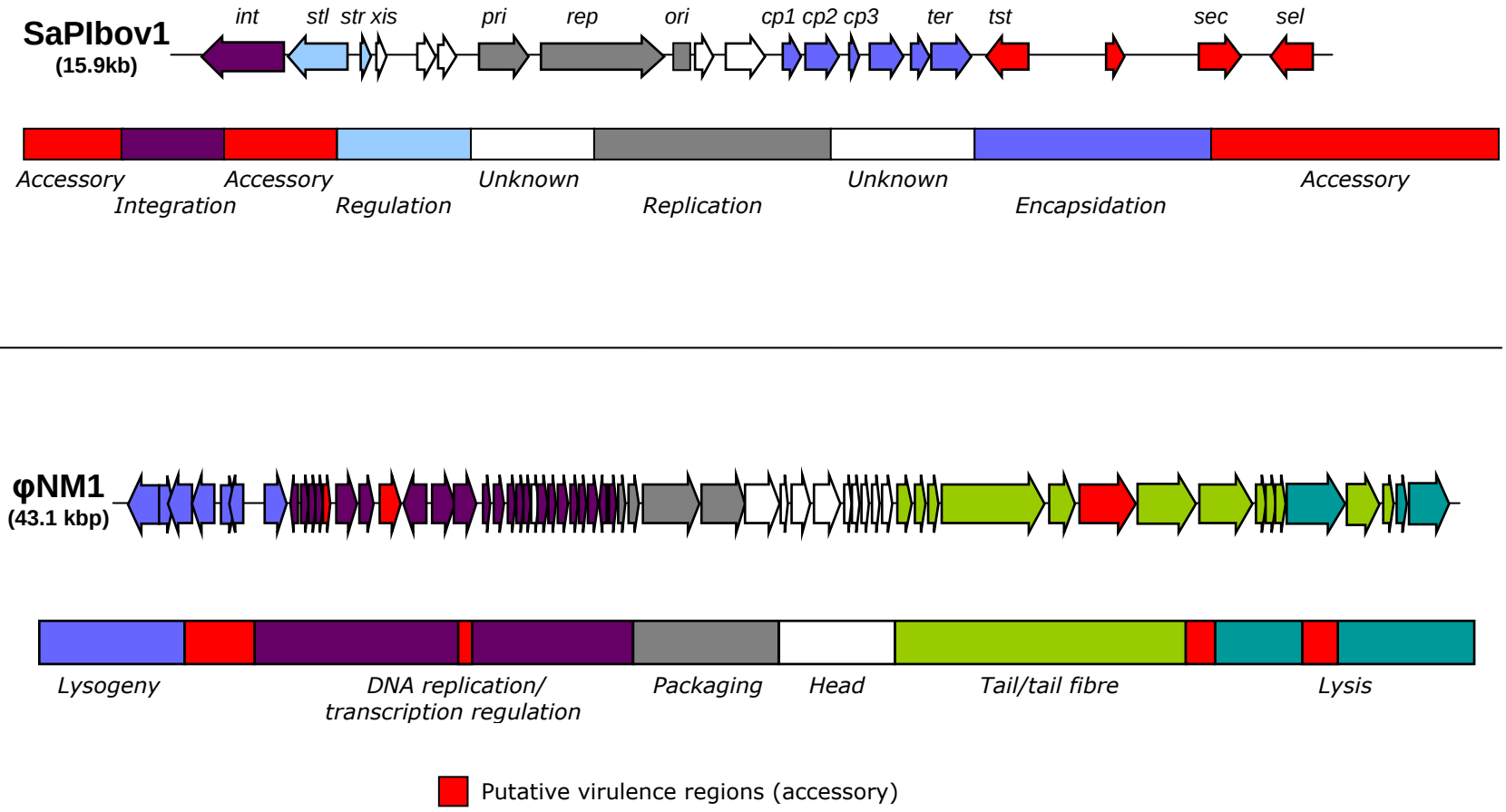


Figure 5. MGE of *S. aureus* exhibit conserved mosaic structure (A) Open reading frames of the pathogenicity Island SaPIbov1. Coloured bars below SaPIbov1 indicate conserved structure of SaPIs. **(B)** Open reading frames of prophage φNM1. Coloured bars below φNM1 indicate conserved structure of the *Siphoviridae* temperate, tailed bacterial viruses most common in *S. aureus*.

S. aureus genomes sequenced to date, and several specific chromosomal integration sites have been identified. Excision and replication is induced by a co-replicating phage, and the element is packaged into phage-like particles and transmitted at a very high frequency (Lindsay *et al.*, 1998, Ruzin *et al.*, 2001, Tormo *et al.*, 2008, Úbeda *et al.*, 2005). The SaPIs, similar to many bacteriophage, have a mosaic structure but a conserved modular organisation (Fig. 5). Each contains a site-specific integrase gene adjacent to two genes encoding putative transcriptional regulators (Subedi *et al.*, 2007), a replication module comprising proteins with helicase and primase activity (Úbeda *et al.*, 2007a) and a six gene encapsidation module (Úbeda *et al.*, 2007b). SaPIs also have three variable regions which often encode virulence factors (Subedi *et al.*, 2007; Fig. 5) including superantigens such as the toxic shock syndrome toxin TSST-1, SelK and SelQ (Fitzgerald *et al.*, 2001a, Úbeda *et al.*, 2003, Yarwood *et al.*, 2002), and the biofilm associated protein Bap (Cucarella *et al.*, 2001).

Plasmids

Plasmids are circular DNA molecules that replicate independently of the main bacterial chromosome. They are frequently transferred between cells, though this transfer system can be blocked in the laboratory using the *Sau*1 restriction modification system (Waldron & Lindsay, 2006). The number of plasmids associated with a strain of *S. aureus* varies widely, and microarray studies often fail to detect plasmid genes, suggesting that substantial sequence variation exists (Lindsay, 2008). Isolates may have multiple plasmids, varying in size from 3 kbp to 150 kbp, which often carry genes involved in resistance to antibiotics, antiseptics and heavy metals (Lindsay, 2008). There are three families, classified by size and ability to conjugate. Class I are the smallest, usually less than 5 kbp, have a high copy number, and can carry one or two antibiotic resistance genes (Khan, 2005). They are sometimes integrated into the chromosome. Class II plasmids are larger, up to 40 kbp, and many confer resistance to β-lactams, heavy metals, antiseptics or aminoglycosides (Lindsay & Holden, 2006). Resistance genes are often part of transposons that have integrated into the plasmid. Class III plasmids are the largest (up to 60 kbp) and are the only conjugative plasmids

due to the presence of *tra* genes (Lindsay & Holden, 2006). They are similar to Class II plasmids with respect to resistance genes carried.

Staphylococcal cassette chromosomes (SCC)

SCCs are integrated into the chromosome at the *orfX* gene, approximately 25 kbp from the origin of replication, and range in size from 3 kbp to 60 kbp (Ito *et al.*, 2001). Several different SCC have been identified encoding genes for resistance to methicillin (SCC*mec*) (Ito *et al.*, 1999, Katayama *et al.*, 2000) or fusidic acid (Lannergard *et al.*, 2009), or those that confer a mucoid capsule type (Luong *et al.*, 2002). The mechanism of HGT between strains in unknown, though it has been suggested that the smaller elements (<45 kbp) transfer by localised transduction (Lindsay, 2008).

Variable genomic regions

vSaα and *vSaβ* are genomic islands identified in *S. aureus* which vary considerably between clonal lineages (Lindsay *et al.*, 2006). There is no evidence that these genomic islands are mobile and they do not typically contain genes involved in mobility (Lindsay, 2008). *vSaα* encodes a cluster of superantigen-like proteins (*ssl*) and ten lipoproteins (Lindsay, 2008). *vSaβ* encodes the *egc* cluster of superantigens, the bi component leukotoxin *lukD* and *lukE*, and the *spl* cluster of serine proteases (Lindsay, 2008). Both islands also encode HsdS and HsdM restriction modification proteins and putative transposases, which in some cases are frameshifted suggesting that they are no longer functional (Lindsay, 2008).

Other MGE

The *S. aureus* genome also contains transposons such as Tn*554*, which is integrated within the SCC*mec* and carries the *ermA* resistance gene (Ito *et al.*, 1999, Katayama *et al.*, 2000). Transposons can also transfer from distinct species, such as the *vanA* transposon that originates with the enterococci (Clark *et al.*, 2005). Multiple copies of transposons may be identified, as they often integrate into the genome in a non-site specific manner. Insertion sequences (IS) are transposons that only encode the transposase, and several types have also been identified in the *S. aureus* genome

(Lindsay, 2008). Multiple copies of the same IS are commonly inserted into the genome at different locations, and may interrupt genes or integrate within promoter regions, affecting transcription and translation of proteins (Lindsay, 2008).

1.4.2. Evolution of the *S. aureus* genome

Microbial evolution often involves adaptation to a variable or novel niche, and can lead to genetic divergence. A niche may consist of a specific environment such as host taxon or tissue type, a particular mode of transmission or intermediate host, or a combination of several such variables (Maiden & Urwin, 2006, Stearns & Hoekstra, 2002). Selection for an advantageous phenotype in a pathogen may result in increased virulence and pathogenicity, causing novel or more severe disease symptoms in the host. Furthermore, new mechanisms for transmission or invasion of host tissue may lead to disease emergence in an entirely new demographic (Maiden & Urwin, 2006).

Analysis of whole genome sequences from different strains can provide insight into the molecular evolutionary history of a species. Change in genome structure and content can occur due to mutation of individual nucleotides, homologous recombination, slippage during replication, chromosome rearrangement events such as inversions, or HGT of mobile genetic elements (Lawrence, 2006, Vos, 2009).

1.4.2.1. Evolution of the accessory genome by HGT

HGT is prevalent in many bacterial species, comprising a key component of the adaptive process as it introduces novel genes faster than may otherwise occur via mutation and/or rearrangement of the existing genome (Ochman *et al.*, 2000). For example, a highly mosaic genome structure in *Pseudomonas aeruginosa* allows different strains to specialise in very different environmental reservoirs (Mathee *et al.*, 2008). Strains effectively customise their genome with acquisition or loss of genomic segments depending on their niche (Mathee *et al.*, 2008). HGT contributes to evolution of the *S. aureus* genome, though the frequency of transfer and the nature of regions transferred varies considerably between lineages.

MGE can be considered 'selfish' in that they utilise the host cell to generate more copies of themselves. Accordingly, they frequently carry genes that provide a selective advantage so that they have a greater chance of being retained (Ochman *et al.*, 2000). These genes may be important for virulence, niche specificity, resistance to antibiotics, regulation, or a particular metabolic pathway (Ochman *et al.*, 2000). A number of virulence and resistance determinants are found on phage, SaPI, plasmids and other MGE in *S. aureus* (Table 2).

Comparison of the genomes of community-acquired and hospital-acquired MRSA strains revealed the effect that different selective pressures can have on HGT. It is suggested that constant exposure to antiseptics and antibiotics in a hospital setting selects for rapid acquisition of MGE, whereas sporadic exposure in the community selects for strains with a simpler, less cumbersome version of SCC*mec* (Baba *et al.*, 2002). *S. aureus* has also been compared to other species of low-GC content Gram positive bacteria (Gill *et al.*, 2005) and coagulase-negative staphylococci (An Diep *et al.*, 2006), revealing the importance of HGT between species for niche adaptation and evolution of pathogenicity.

Selective pressure for an HGT event may be sufficiently great that transfer occurs from distant taxa, despite barriers that exist to minimise this. A notable example is the transfer of Tn*1546* carrying *vanA* from *Enterococcus faecalis* to *S. aureus*, conferring resistance to vancomycin (Chang *et al.*, 2003, Weigel *et al.*, 2003). In the case of the first reported VRSA isolate, resistant strains of both species were isolated from a catheter tip, suggesting that transfer occurred *in vivo* (Chang *et al.*, 2003, Weigel *et al.*, 2003). Furthermore, the patient had been treated with vancomycin for several weeks due to recurrent MRSA infection (Chang *et al.*, 2003), providing the selective pressure for acquisition of an MGE carrying resistance genes.

Table 2. Mobile genetic elements (MGE) in *S. aureus*. Examples of MGE, virulence genes and antibiotic resistance determinants encoded, and their role in pathogenesis.

Mobile Genetic Element		Virulence factors / resistance determinants	Role in pathogenesis	Refs
Phage	β-converting phage	Immune Evasion Cluster: • Staphylokinase (*sak*) • Staphylococcal enterotoxin A (*sea*) • Staphylococcal enterotoxin P (*sep*) • Chemotaxis inhibitory protein of *S. aureus* (CHIPS) • Staphylococcal complement inhibitor (SCIN)	Evasion of the human immune system.	van Wamel *et al.*, 2006
	φSLT / φPVL	Panton-Valentine Leukocidin (lukS-PV and lukF-PV)	Leukocytolysis and tissue necrosis.	Narita *et al.*, 2001
	φETA	Exfoliative toxin A (*eta*)	Dual toxin. Induces exfoliation by displaying serine protease activity. Displays superantigenic activity.	Yamaguchi *et al.*, 2000 Vath *et al.*, 1997

Mobile Genetic Element		Virulence factors / resistance determinants	Role in pathogenesis	Refs
SaPI	SaPIbov1	• Toxic shock syndrome toxin (*tst*) • Staphylococcal enterotoxin L (*sel*) • Staphylococcal enterotoxin C (*sec*)	Superantigens. Modulate host immune response, may contribute to evasion of host defences and bacterial persistence.	Fitzgerald *et al.*, 2001a
	SaPI3	• Ear (*sapi3_1 ear*) • Staphylococcal enterotoxin K (*sek*) • Staphylococcal enterotoxin Q (*seq*)	Ear: Putative β-lactamase activity. SEK/Q: Superantigens (see above).	Yarwood *et al.*, 2002
	SaPIbov2	• Biofilm associated protein Bap (*bap*) (on transposon-like elements integrated into SaPI)	Enhances persistence in sheep intramammary and mouse infection models.	Úbeda *et al.*, 2003 Cucarella *et al.*, 2001
Plasmid	pT181	• Teracycline resistance protein (*tet*)	Tetracycline resistance.	Gill *et al.*, 2005 Khan & Novick, 1983
	pETB	• Exfoliative toxin B	Induces exfoliation by displaying serine protease activity.	Yamaguchi *et al.*, 2001
SCC	SCC*mec*	• Penicillin-binding protein 2' (*mecA*)	Methicillin resistance.	Katayama *et al.*, 2000

Variation in restriction modification genes can also influence frequency of HGT. In addition to the lineage-specific differences discussed above, evolution of a single clone can be shaped by features of its restriction modification system. For example, strains of the bovine associated lineage ST151 have mutations in both copies of *sau*1*hsdS* and are hyper-susceptible to uptake of DNA from distant taxa by conjugation (Sung & Lindsay, 2007). In a model of *vanA* transfer, ST151 isolates were 500 times more likely to become recipients of the enterococcal DNA than human associated lineages (Sung & Lindsay, 2007). Furthermore, analysis of the ST151 genome reveals extensive diversification, due in part to the acquisition of MGE encoding toxins not made by closely related strains (Guinane *et al.*, 2008)

1.4.2.2. Genome evolution by gene loss

Several studies have highlighted the importance of genome reduction in bacteria, previously assumed to contain relatively little 'junk' DNA in comparison to eukaryotic genomes (Liu *et al.*, 2004). Insertion or deletion of one or small number of bases can lead to a frameshift in a gene, or substitution can lead to a premature stop codon, both of which inactivate the gene forming a pseudogene (Lerat & Ochman, 2005). In some cases, genes are disrupted by integration of an IS element, a process that appears to be more frequent in some species that others (Lerat & Ochman, 2005). Pseudogenes can be identified by genome comparisons which detect ORFs that differ significantly in length to homologues in closely related strains or species.

An analysis of 64 phylogenetically diverse prokaryote genomes identified approximately 7,000 candidate pseudogenes, which accounts for 1 % to 5 % of all gene-like sequences (Liu *et al.*, 2004). The proportion of the genes found to be inactive varied greatly, with some species having considerably higher occurrence (Liu *et al.*, 2004). Interestingly, pseudogenes are more than twice as likely as genes to have anomalous codon usage associated with horizontal transfer. This suggests that gene loss may be important for eliminating genes acquired by HGT, if they are deleterious or neutral with respect to host fitness. A separate study also considered a range of species, but in this case some

were closely related and thus had broadly overlapping gene inventories (Lerat & Ochman, 2005). Nonetheless, each genome revealed a largely unique set of pseudogenes, suggesting that they are formed and eliminated rapidly from most bacterial genomes (Lerat & Ochman, 2005). Pseudogenes represent <3 % of the genome in *S. aureus*, a relatively low figure compared to some other bacterial species though not markedly different from the average (Lerat & Ochman, 2005, Liu *et al.*, 2004).

Whole genome sequencing projects have been performed for numerous bacterial species, and in several cases have highlighted a role for pseudogene formation in niche adaptation. Fragmented genes are prevalent in strains of *Helicobacter pylori* and *Salmonella enterica* that have undergone a host switch, or changed from host promiscuity to infection of a single host species (Eppinger *et al.*, 2006, Thomson *et al.*, 2008). They are also common in lactic acid bacteria that have been restricted to a particular environmental niche by human intervention, such as for yoghurt production (van de Guchte *et al.*, 2006). Extreme examples of genomic erosion are found in obligate intracellular pathogens such as the insect symbiont *Buchnera aphidicola* (Moran *et al.*, 2009) and human pathogen *Mycobacteium leprae* (Cole *et al.*, 2001).

1.4.2.3. Genome evolution by positive selection

Positive selection leading to fixation of a beneficial mutation is an important facet of microbial adaptation, and can play a key role in niche adaptation. In recent years, phylogenetic methods have been applied to many taxa to identify genes under positive selection by comparing rates of non-synonymous versus synonymous mutation (dN/dS). A dN/dS ratio greater than one indicates that, for a given gene or gene fragment, most of the nucleotide changes will affect the amino acid sequence of the encoded protein. This means that there is more likely to be a phenotypic effect, suggesting that positive selection has been acting upon the sequence to maintain these mutations as they derive some benefit (Maiden & Urwin, 2006). Conversely, synonymous substitutions are assumed to have no effect on the fitness of the organism as they do not alter the protein. Therefore, if the rate of non-synonymous substitutions is comparable to that of

synonymous substitutions, the DNA sequence is evolving neutrally. Finally, a dN/dS ratio below one indicates purifying selection, whereby non-synonymous substitutions are removed from the population in order to preserve the function of the protein.

Early models for calculating dN/dS to infer positive selection were sensitive to heterogeneity across the genome with respect to rates of recombination, G+C content and codon usage (Hartl & Clark, 2007). Improved models have been developed to account for such variation, most notably codon-substitution models that take into account all possible combinations of codons and track each codon separately (Nielsen & Yang, 1998), and models that allow for rate variation between lineages (Yang & Nielsen, 2002, Yang *et al.*, 2005). Such models estimate the parameter ω, defined as the relative rate of non-synonymous to synonymous substitutions (Hartl & Clark, 2007).

dN/dS (or ω) analysis has been performed with several bacterial species on a genome-wide scale, considering all genes rather than those whose significance has been recognised previously. It is important to identify genes under positive selection, as this could lead to enhanced understanding of the mechanisms of bacterial adaptation and pathogenesis, and aid in the development of novel therapeutics. Such a study has not been performed with *S. aureus*, but analyses of several other species have generated interesting results. For example, a comparative study of *E. coli* and its close relative *Shigella* revealed positive selection acting across all lineages, affecting surface proteins such as beta barrel porins and, unexpectedly, transposons (Petersen *et al.*, 2007). It was suggested that selection acting upon transposons relates to genomic conflict between host bacteria and transposable elements (Petersen *et al.*, 2007). A separate study of *E. coli* searched for genes under selection in the uropathogenic lineage only, with the aim of identifying genes important for persistence and pathogenesis in the urinary tract environment (Chen *et al.*, 2006a). Genes identified encode proteins involved in cell surface structure, DNA metabolism, nutrient acquisition, and establishment of urinary tract infection (Chen *et al.*, 2006a).

Rates of positive selection vary between different bacterial taxa, due to contrasting selective pressures and evolutionary history. In *Streptococcus*, positive selection is observed in 11 % to 34 % of the core genome and has an important role in species differentiation and host adaptation. However, it is a slow process and has had less of an impact on genome structure than recombination (Lefebure & Stanhope, 2007). By contrast, 92.5 % of the nonrecombinant core genome of *Campylobacter* is under positive selection in at least one lineage (Lefebure & Stanhope, 2009). This has been attributed to intensive interspecific competition for resources in the gastrointestinal tract inhabited by *Campylobacter*, resulting in a 'Red Queen Hypothesis' evolutionary scenario whereby each competing resident must contantly adapt in order to maintain its position within the ecosystem (Lefebure & Stanhope, 2009, van Valen, 1973).

Variation has also been observed between different clades of *Listeria* that differ in breadth of environmental range. Higher rates of both positive selection and recombination are observed in lineages with a broader niche range than in those that have become specialised (Orsi *et al.*, 2008). A comparison of *S. enterica* serotypes revealed similar patterns, in as much as branch specific positive selection contributes to the evolution of different lineages (Soyer *et al.*, 2009). Like the *Listeria* clades, lineages differ in their degree of host-restriction. Such findings highlight the importance of positive selection in niche adaptation, including host specialisation.

As many of these studies have shown, positive selection analysis can detect specific genes important for bacterial persistence or pathogenesis in a given niche, often encoding proteins at the interface between host and pathogen, or those essential for unique metabolic requirements (Chen *et al.*, 2006b, Petersen *et al.*, 2007). Further analysis can be performed to highlight the residues targeted by positive selection, gaining insight into the role of particular regions in protein function. Petersen *et al.* (2007) found that extracellular regions of cell surface proteins made by *E. coli* and *Shigella flexneri* show evidence for positive selection, presumed to be due to their role in interaction both with the host and with competing microbes.

Scrutiny of gene sequences for evidence of positive selection has revealed which genes are important to pathogens inhabiting a range of different niches, and shed light on how microbial proteins may interact with the host. Furthermore, genome-wide analysis has revealed the relative contribution of positive selection to genome evolution in pathogens with differing lifestyles. It is clear that such an approach provides a valuable tool for understanding the evolutionary history of a bacterial species, and may emphasise the importance of hitherto unknown genes in disease pathogenesis.

Of note, *Rocha et al.* have demonstrated that the dN/dS ratio is not static but decreases over time due to gradual selective purging of slightly deleterious (nonsynonymous) mutations. The trajectory of the slope of this decrease is governed by the strength of purifying selection, which in turn depends on the selective coefficient and also the effective population size (Ohta *et al.*, 1973, Rocha *et al.*, 2006). This subject was further explored in a study of *E.coli* and *Shigella* genomes, in which the authors applied a time-dependant approach to identify evidence for reduced purifying selection within the latter, compared with the former (Balbi *et al.*, 2008). They revealed that frequencies of various polymorphism types change in a predictable manner over time as certain types are preferentially eliminated over their counterparts, again at a rate determined, at least in part, by selective coefficient and effective population size (Balbi *et al.*, 2008). Taken together, this suggests that particular genomic changes (and thus altered dN/dS ratio) may be consistent with relaxed or inefficient selection owing to a reduced population size in a particular evolutionary lineage. This should be considered when interpreting the results of dN/dS analyses.

1.4.2.3.1. Methods for the identification of positive selection

Phylogenetic Analysis by Maximum Likelihood (PAML) is a package of programmes that implement a wide variety of models to explore the evolutionary relationships between sequences at either the protein, codon or DNA level, using maximum likelihood. PAML allows the user to discover more about the phylogenetic evolutionary model that has shaped a group of organisms, via analysis of a representative set of

sequences. This is a widely used package, generally recognised as the gold standard for such studies.

The two basic components of a phylogenetic evolutionary model are a phylogeny (a phylogenetic tree) and an evolutionary model (a model of how sequences evolve by substitutions and other events over time). Phylogeny can be calculated using separate, dedicated software such as HyPhy (Pond *et al.*, 2005) or Molecular Evolutionary Genetic Analysis (MEGA; Tamura *et al.*, 2007), and specified in the input for PAML. PAML then performs analyses based upon rate matrices, which describe the rate at which one character (nucleotide, amino acid, or codon) is replaced by another. Each row of a rate matrix must equal zero, so the rate at which a character does not change is minus the sum of the rate at which it does change into another character. The programme identifies the evolutionary model and substitution rate that maximises the likelihood of observing the input sequences, i.e. selects the model that best fits the data.

This process can be used to perform several analyses, including estimation of nonsynonymous and synonymous mutation rates (dN and dS and dN/dS) performed by the codeml programme. This can be applied to test hypotheses about rates of evolution, including the presence of positive selection.

1.5. Human activity and pathogen evolution

Human activity that alters the environment inhabited by a microbial pathogen, either directly or indirectly, can also have major consequences for pathogen evolution (Lebarbenchon *et al.*, 2008). Our actions can lead to the emergence of microorganisms with novel transmission mechanisms and phenotypic traits, as they adapt to exploit a novel niche. Factors ranging from habitat disruption, pollution and climate change to agricultural practice and medical treatment strategies can have a profound effect on host-microbe interactions (for review see Lebarbenchon *et al.*, 2008). For example, domestication of livestock, which began around 10,000 years ago, and the intensive

farming practices of the last century, have drastically altered ecosystem dynamics. In addition, urbanisation and globalisation can have a major impact on pathogen evolution.

1.5.1 The rise of agriculture

For most of human evolutionary history, populations have existed as isolated hunter-gatherer communities with limited contact between groups. However, a transition from hunter-gatherer to sedentary settlements, which would eventually become the urbanisations of the modern world, was brought about by domestication of animals and plants to provide a more reliable and less labour-intensive food source. This began around 10,000 years ago when wild animals were tamed and utilised for their meat, for other forms of produce such as hide or dairy products, or for manual labour (Clutton-Brock, 1981, Diamond, 2002). The new way of life resulted not only in higher population densities, but also co-habitation of humans and domesticated livestock. This was accompanied by an increase in zoonotic and anthroponotic pathogens as humans and animals were in close, direct and prolonged contact (for review see Pearce-Duvet, 2006). Furthermore, the new host population structure allowed for microbes with altered transmission mechanisms and an increase in severity of host damage (i.e. disease symptoms), leading to the emergence of so-called crowd diseases.

Before domestication, some species of wild animal would also reside in small, heterogeneous populations, forming part of a dynamic ecosystem in which long term evolution has shaped the balance between component species. Such species are thus susceptible to infection by novel microbes when forced to live in large homogeneous groups for farming purposes. Other species, however, naturally live in large social groups which can support crowd diseases (Clutton-Brock, 1981), and are in fact predicted to be the source of zoonoses that switch host and emerge as human pathogens (Fiennes, 1978, McNeill, 1989, Pearce-Duvet, 2006).

1.5.2. The modern world

In recent decades, prominent examples have come to the fore in which agricultural practices have imposed selective pressures leading to different mechanisms for microbial transmission and opportunities for emergence in a novel demographic. Notably, the bovine spongiform encephalopathy (BSE) epidemic of the 1990s was due largely to the use of meat and bone meal in cattle feed, and was compounded by a high sheep to cattle ratio in the UK farming industry (Nathanson *et al.*, 1997). In addition, during the foot and mouth disease epidemic of 2001, spread of the disease was enhanced by the long distance transport of livestock (Ferguson *et al.*, 2001).

Globalisation and the growth of international air travel have also had a major impact on pathogen transmission mechanisms. Modern human lifestyles present few barriers to microbes, and emerging pathogens are able to spread rapidly between countries. Intercontinental travel has been implicated in the spread of acute haemorrhagic conjunctivitis in the early 1980s (Kew *et al.*, 1983) and 2000s (Dussart *et al.*, 2005), of *Neisseria meningitidis* associated with the Hajj pilgrimage (Wilder-Smith & Memish, 2003), and of the Severe Acute Respiratory Syndrome (SARS) virus in the 2002-2003 outbreak (Knobler *et al.*, 2004; Fig. 6).

1.6. Applications of molecular genetic analysis of pathogens

In order to develop effective prevention strategies and therapeutics for infectious diseases, it is important to understand the evolutionary processes and molecular mechanisms that ultimately shape a virulent pathogen. Knowledge of the epidemiology and population genetics of a microbe can greatly improve contingency plans, making society better prepared for major outbreaks when they occur. In the case of SARS, a potentially devastating pandemic was successfully contained in less than 4 months, due primarily to an unprecedented level of international cooperation, and to planning based

on sound scientific knowledge (Knobler *et al.*, 2004). In another notable example, it is suggested that reconstruction of the global picture of annual strain emergence of influenza A virus (Rambaut *et al.*, 2008, Russell *et al.*, 2008) will enhance effective vaccine selection.

Identification of biological processes, individual proteins and amino acid residues that play a crucial role in bacterial infection can lead to development of effective therapeutics. Genome scale analysis can be used to identify potential targets for antimicrobial pharmaceuticals (Shen *et al.*, 2010). Furthermore, thorough understanding of the molecular basis of pathogenicity has been exploited to develop drugs that disrupt essential processes such as quorum-sensing (Hentzer *et al.*, 2003), rather than killing the microbe. Promising results have also been seen by impeding

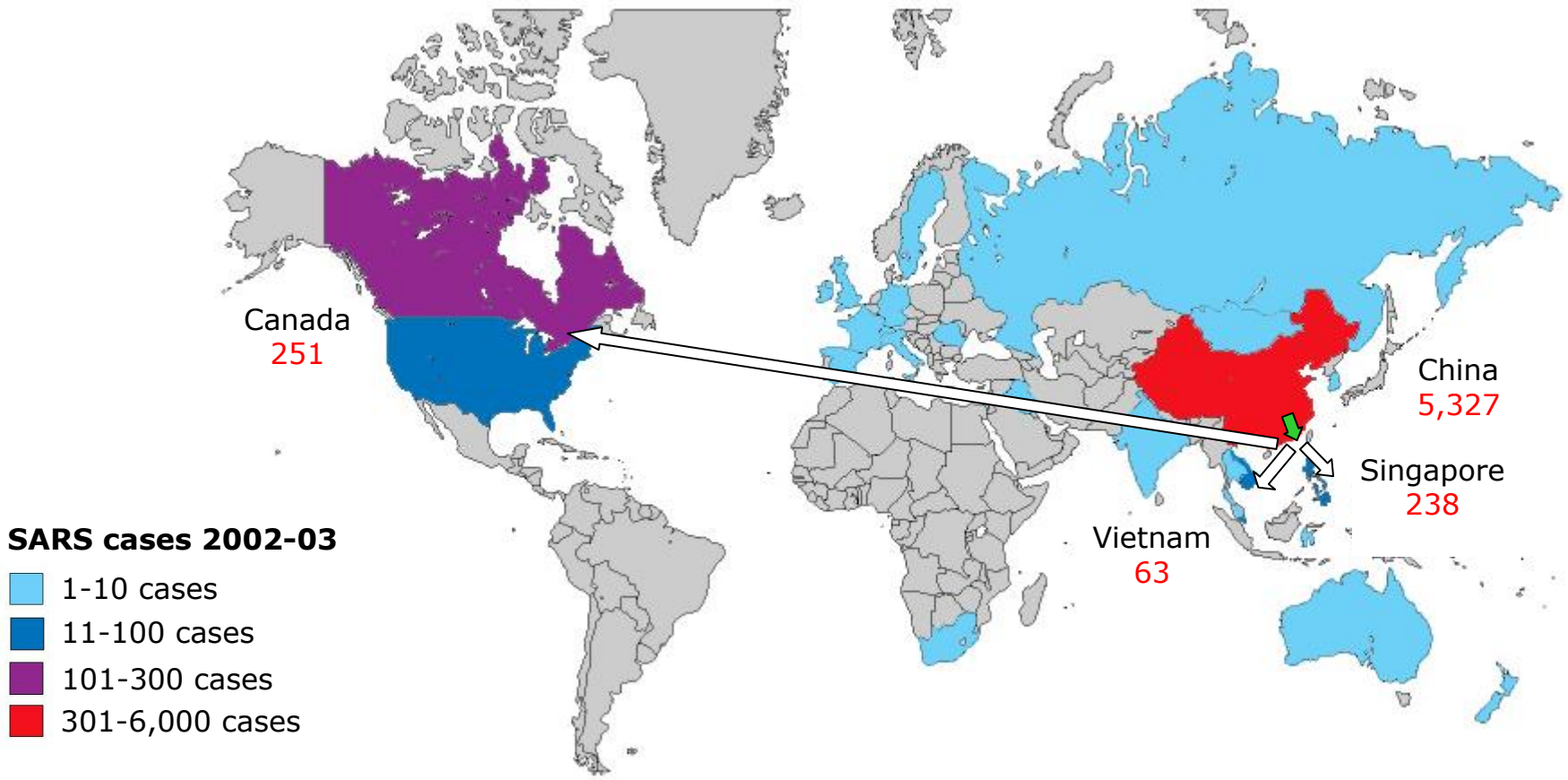


Figure 6. Global air travel can enhance rapid spread of infectious disease agents. Severe Acute Respiratory Syndrome (SARS) originated in Guandong province, China. Virus spread to Hong Kong (green arrow) and from there to Singapore, Hanoi and Toronto (white arrows) with airline passengers. Cases were then identified across the globe. Colours indicate probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 (data published by the World Health Organisation at http://www.who.int/csr/sars/country/table2003_09_23/en/index.html).

mechanisms of persistence and adaptation in bacteria, an approach that could be used to reduce the likelihood of bacteria evolving resistance to new antibiotics (Smith & Romesberg, 2007). Vaccine discovery also benefits from a deeper understanding of both molecular biology and population genetics. A comparative genomic approach termed reverse vaccinology has proven successful for identifying candidate surface proteins, whilst avoiding problems of excessive variation between different strains and populations. The technique has been applied to several species including *Neisseria meningitidis* B (Giuliani *et al.*, 2006), group B *Streptococcus* (Maione *et al.*, 2005) and *S. aureus* (Stranger-Jones *et al.*, 2006) and will no doubt enhance our ability to prevent infectious diseases over the coming decades.

1.7. Aims of the study

- To characterise the genetic diversity of the global population of poultry isolates of the major bacterial pathogen *S. aureus.*

- To establish the phylogenetic relationships between *S. aureus* strains isolated from poultry and those from distinct avian species and other host taxa.

- To determine the genomic basis for host specificity of poultry associated *S. aureus,* and to compare and contrast with that of strains from close or distantly related lineages.

- To assess the frequency and importance of positive selection in evolution of the *S. aureus* species, and in particular for host adaptation.

2. Materials and methods

2.1. Bacterial growth conditions

Bacteria were grown to stationary phase in tryptic soy broth (TSB) or brain and heart infusion (BHI) broth at 37 ºC for 16 h with shaking at 200 rpm. For exponential phase, 200 µl of stationary phase culture was added to 40 ml fresh TSB and grown to an OD_{600} value of 0.6. For short term storage and isolation of individual colonies, culture was streaked out on tryptic soy agar (TSA) plates, incubated at 37 ºC overnight and stored at 4 ºC. Bacterial culture was stored long term as glycerol stock at -80 ºC.

2.2. Genomic DNA extraction

DNA was isolated from 1 ml volume of stationary phase *S. aureus* TSB culture using the Edge Biosystems Bacterial Genomic DNA Purification Kit (Edge Biosystems, MD, USA) according to the manufacturers instructions with addition of lysostaphin (AMBI Products LLC, NY, USA) (5 mg ml^{-1}) to the cell lysis step.

2.3. Polymerase Chain Reaction (PCR)

PCR reactions were performed for detection of particular regions in a given strain, or to generate template for directed sequencing reactions. Reactions were performed in a Biometra TGradient thermocycler.

Standard PCR reactions were performed in 50 µl reaction volumes containing approximately 10 ng of genomic DNA, 200 nM of each primer, 1 U of *GoTaq* DNA polymerase (Invitrogen, Paisley, UK), 1 X PCR buffer, 1.5 mM $MgCl_2$ (Promega, Hampshire, UK) and 0.2 mM deoxynucleoside triphosphates (Promega, Hampshire, UK). Reaction conditions were 94 °C for 3 min, followed by 30 cycles of 94 °C for 0.5 min to 1 min, 42 °C to 65 °C (depending upon primer annealing temperature) for 0.5 min to 1 min and 72 °C for 1 min/kbp, followed by 72 °C for 5 min.

PCR reactions requiring a high fidelity, long-range polymerase were performed in 50 µl reaction volumes, containing approximately 10 ng of genomic DNA, 45 µl Platinum

PCR Supermix (Invitrogen, Paisley, UK) and 200 nM of each primer. Reaction conditions were 94 °C for 0.5 min to 2 min, followed by 25 to 35 cycles of 94 °C for 0.25 min to 0.5 min, 55 °C for 0.25 to 0.5 min and 68 °C for 1 min per kbp.

Alternatively, the 50 µl reaction volume contained approximately 10 ng of genomic DNA, 200 nM of each primer, 1 U of PFU Ultra II Fusion HS DNA Polymerase (Promega, Hampshire, UK), 1 X PCR buffer and 0.2 mM deoxynucleoside triphosphates (dNTPs) (Promega, Hampshire, UK). For products <10 kbp reaction conditions were 94 °C for 2 min, followed by 30 cycles of 94 °C for 20 s, 42 °C to 65°C (depending upon primer annealing temperature) for 20 sec and 72 °C for 15 s per kbp, followed by 72 °C for 3 min. For products >10 kbp reaction conditions were 92 °C for 2 min followed by 30 cycles of 92 °C for 10 s, 42 °C to 65 °C (depending upon primer annealing temperature) for 20 s and 68 °C for 30 s per kbp, followed by 68 °C for 5 min.

Products were resolved by electrophoresis in a 1 % (w/v) agarose gel, and visualised using 0.003 % (w/v) Ethidium Bromide or 0.007 % (w/v) Gel Red (Biotium) stains with an ultraviolet transilluminator (MultiImage™ Light Cabinet, Alpha Innotech Corporation).

2.4. PCR product purification

PCR products were precipitated using one of three methods.

(i) QIAquick PCR Purification Kit (Qiagen), following supplied protocol.

(ii) Polyethylene Glycol (PEG) / NaCl precipitation.
Each 50 µl PCR reaction was incubated with 60 µl 20 % (w/v) PEG/2.5M NaCl for 14 h at 4 ºC. Products were pelleted by centrifugation at 2750 g for 1 h at 4 ºC, and supernatant removed by centrifugation inverted at 500 x *g* for 1 min. Pellets were washed twice with ice-cold 70 % (w/v) ethanol, by centrifugation at 2750 x *g*

for 10 min followed by centrifugation inverted at 500 x *g* for 1 min. Products were re-suspended in 50 µl dH_2O.

(iii) Exonuclease I-Antarctic Phosphatase (Exo-AP) digestion.
In a 10 µl reaction, 4 µl PCR product was combined with 1 X Antarctic Phosphatase buffer (New England Biolabs), 100 U Exonuclease I (NEB) and 100 U Antarctic Phosphatase (NEB). Reaction was incubated at 37 ºC for 15 min, 80 ºC for 15 min, and subsequently held at 15 ºC, using a Biometra TGradient or Bio-Rad MyCycler thermocyling machine.

2.5. Directed DNA sequencing

Directed sequencing was performed by The Gene Pool at the University of Edinburgh (http://genepool.bio.ed.ac.uk). DNA sequence extension products were generated from purified PCR products using Applied Biosystems BigDye Terminator Cycle Sequencing protocol, and resolved by capillary electrophoresis with a 3730 instrument (Applied Biosystems). Chromatograms were visualized and edited using Pregap4 and Gap4 (Staden *et al.*, 1998), Vector NTI (Invitrogen, Paisley, UK) or Geneious (Biomatters Ltd.) software packages.

2.6. Phylogenetic analysis

Reconstruction of evolutionary relationships was carried out using the MEGA 4 package (Tamura *et al.*, 2007). Trees were constructed using the neighbor-joining method (Saitou & Nei, 1987) with bootstrapping consensus inferred from 500 replicates (Felsenstein, 1985). Where applicable, the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to branches.

3. Population genetic analysis of avian associated *S. aureus*

3.1. Introduction

Population genetic analyses of *S. aureus* have revealed a great deal about diversity within the species. Evolution of separate lineages has been largely clonal, with low rates of homologous recombination (Feil *et al.*, 2003, Lindsay *et al.*, 2006). Geographic variation has been observed as some clades are responsible for disease across a broad area, whereas others have been identified in only a small number of locations (Fan *et al.*, 2009, Feil *et al.*, 2003, Melles *et al.*, 2008, Ruimy *et al.*, 2009). Further, population genetic studies have shed light on the evolutionary history of clinically important clones. It has been suggested that MRSA is the result of recurrent HGT of SCC*mec* into strains of multiple distinct clonal lineages as opposed to dissemination of a single MRSA clone (Enright *et al.*, 2002, Fitzgerald *et al.*, 2001b, Musser & Kapur, 1992), and that commensal and invasive isolates have the same clonal origins supporting the hypothesis that *S. aureus* is an opportunistic pathogen (Feil *et al.*, 2003, Lindsay *et al.*, 2006).

However, studies have focussed primarily on human isolates, with most studies that consider animal strains including those of cattle or other ruminants. Where other animal isolates have been analysed they represent a limited sub-set of the global population. It is important that analysis of large collections of *S. aureus* isolates from all major livestock species is performed. This may establish the existence of host-specific clones, further scrutiny of which could reveal unique genotypic and/or phenotypic characteristics, leading to enhanced understanding of the processes involved in bacterial evolution and niche adaptation. Alternatively, studies may reveal host-promiscuity, a trait with important implications for potential zoonotic or anthroponotic transfer and pathogen emergence. The aim of this study was to investigate *S. aureus* isolates from broiler chickens intensively farmed in geographically distant countries, using MLST.

3.2. Aims

- To determine population structure of poultry associated strains of *S. aureus* isolated from broiler chickens on farms in several countries.
- To ascertain the existence of population structure according to geographical location, as has been observed in human isolates.
- To determine the phylogenetic relationship between *S. aureus* poultry isolates and those of other avian species.

3.3. Materials and methods

3.3.1. Strains analysed and bacterial growth conditions

Phylogenetic analysis was performed on 48 isolates of *S. aureus* isolated from broiler chickens (*Gallus gallus*) on farms in 6 countries on 4 different continents (Table 3). The strain collection comprised 11 isolates from Northern Ireland (UK), 2 from England (UK), 9 from Georgia (USA), 1 from Iowa (USA), 7 from Belgium, 5 from Australia, 8 from Japan and 5 from Denmark. Of these strains, 39 % were commensal isolates and 58 % were clinical isolates (1 strain was of unknown clinical origin). Strains from England were isolated prior to 1954 and other strains were obtained in the 1970s, 1990s or in the last decade.

In addition, 9 isolates were analysed from different species of reared game and wild birds in Scotland (Table 3) including partridge (*Perdix perdix*), pheasant (*Phasianus colchicus*), a poultry layer (*Gallus gallus*; non-commercial farm), a turkey (*Meleagris gallopavo*) and a wild buzzard (*Buteo buteo*).

3.3.2. Multi Locus Sequence Typing (MLST)

MLST was performed by PCR amplification and sequencing of internal fragments of seven housekeeping genes, *arcc, aroe, glpF, gmk, pta, tpi* and *yqil*. Protocol was as

Table 3. Origin and characteristics of *Staphylococcus aureus* strains examined in this study

Strain *	Host (disease or characteristic)	Geographic origin	Isolation date	Reference	MLST (clonal complex)
ED98 (ch1)	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV72	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV4 (ch2)	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV24 (ch3)	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV35	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV86 (ch17)	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	692 (CC385)
AV136	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV38	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV92	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV70	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
AV39	Broiler Chicken (skeletal infection)	N. Ireland (UK)	1996-97	Rodgers *et al.*, 1999	5 (CC5)
9612 (ch16)	Broiler Chicken (septicaemia/arthritis)	England (UK)	Pre 1954	Smith, 1954	692 (CC385)
9715	Broiler Chicken (septicaemia/arthritis)	England (UK)	Pre 1954	Smith, 1954	8 (CC8)
00-8661	Broiler Chicken (hock isolate)	Georgia (USA)	1999	This study	5 (CC5)
00-7439 (ch9)	Broiler Chicken (hock isolate)	Georgia (USA)	1999	This study	5 (CC5)

Strain *	Host (disease or characteristic)	Geographic origin	Isolation date	Reference	MLST (clonal complex)
99-7184	Broiler Chicken	Georgia (USA)	1999	This study	5 (CC5)
00-9705 (ch8)	Broiler Chicken (synovial fluid isolate)	Georgia (USA)	1999	This study	5 (CC5)
00-8078	Broiler Chicken (hock isolate)	Georgia (USA)	1999	This study	5 (CC5)
10009 (ch10)	Broiler Chicken (sternal bursa isolate)	Iowa (USA)	1999	This study	1342 (CC5)
00-8954	Broiler Chicken (synovial fluid isolate)	Georgia (USA)	1999	This study	5 (CC5)
10217	Broiler Chicken (purulent stifle/hock joints)	Georgia (USA)	2000	This study	5 (CC5)
00-9686	Broiler Chicken (stifle isolate)	Georgia (USA)	1999	This study	5 (CC5)
99-5281	Broiler Chicken (hock isolate)	Georgia (USA)	1999	This study	5 (CC5)
CIX2 (ch3)	Broiler Chicken (commensal)	Belgium	1976	This study	5 (CC5)
CIX18 (ch15)	Broiler Chicken (commensal)	Belgium	1976	This study	385 (CC385)
CIX38 (ch5)	Broiler Chicken (commensal)	Belgium	1976	This study	5 (CC5)
CIX51	Broiler Chicken (commensal)	Belgium	1976	This study	5 (CC5)
CIX8	Broiler Chicken (commensal)	Belgium	1976	This study	385 (CC385)
CIX77	Broiler Chicken (commensal)	Belgium	1976	This study	5 (CC5)
CIX90	Broiler Chicken (commensal)	Belgium	1976	This study	1343 (CC385)

Strain *	Host (disease or characteristic)	Geographic origin	Isolation Date	Reference	MLST (clonal complex)
4P38 (ch14)	Broiler Chicken (commensal)	Australia	1995-97	Bertolatti *et al.*, 2003	8 (CC8)
6P061 (ch12)	Broiler Chicken (commensal)	Australia	1995-97	Bertolatti *et al.*, 2003	1345 (CC1)
7P79	Broiler Chicken (commensal)	Australia	1995-97	Bertolatti *et al.*, 2003	1345 (CC1)
6P070 (ch11)	Broiler Chicken (commensal)	Australia	1995-97	Bertolatti *et al.*, 2003	1345 (CC1)
6P72	Broiler Chicken (commensal)	Australia	1995-97	Bertolatti *et al.*, 2003	1345 (CC1)
CH69 (ch13)	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	1 (CC1)
CH48 (ch19)	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	1027
CH53	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	1027
CH62 (ch20)	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	1027
CH77 (ch6)	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	5 (CC5)
CH38 (ch7)	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	5 (CC5)
CH43	Broiler Chicken (commensal)	Japan	1998	Takeuchi *et al.*, 2002	1350 (CC5)
CH91 (ch18)	Broiler Chicken (edematous & necrotic dermatitis)	Japan	1967	Takeuchi *et al.*, 2002	1344 (CC385)
A8	Broiler Chicken (clinical isolate)	Denmark	2006-07	This study	5 (CC5)
C3	Broiler Chicken (clinical isolate)	Denmark	1998	This study	5 (CC5)

Strain *	Host (disease or characteristic)	Geographic origin	Isolation date	Reference	MLST (clonal complex)
C8	Broiler Chicken (clinical isolate)	Denmark	1998	This study	5 (CC5)
D1	Broiler Chicken (clinical isolate)	Denmark	1998	This study	5 (CC5)
D2	Broiler Chicken (clinical isolate)	Denmark	1998	This study	5 (CC5)
B600539 (ph2)	Farmed pheasant (tenosynovitis)	Scotland (UK)	Unknown	This study	692 (CC385)
B305236/1 (pa3)	Farmed partridge (liver, secondary infection)	Scotland (UK)	2006	This study	692 (CC385)
B306222 (bu1)	Wild buzzard (carpals, secondary infection)	Scotland (UK)	2008	This study	692 (CC385)
B305487 (ch11)	Backyard poultry layer (spleen, secondary infection)	Scotland (UK)	2006	This study	692 (CC385)
B34/96/1 (pa1)	Farmed partridge (clinical isolate)	Scotland (UK)	1996	This study	5 (CC5)
B304223 (tu1)	Farmed turkey (conjunctivitis)	Scotland (UK)	Unknown	This study	5 (CC5)
B95/97/1 (pa2)	Farmed red-legged partridge (clinical isolate)	Scotland (UK)	1997	This study	1346 (CC5)
B600552 (ph1)	Farmed pheasant (tenosynovitis)	Scotland (UK)	Unknown	This study	1347 (CC385)
B600535	Farmed pheasant (tenosynovitis)	Scotland (UK)	Unknown	This study	15 (CC15)

Strain *	Host (disease or characteristic)	Geographic origin	Isolation date	Reference	MLST (clonal complex)
PL72 (hu4)	Human (hospital infection)	Poland	1991	Leski *et al.*, 1998	5 (CC5)
MR1 (hu3)	Human (wound isolate)	Poland	1992	Leski *et al.*, 1998, Trzciński *et al.*, 1997	5 (CC5)
N315 (hu1)	Human (pharyngeal smear)	Japan	1982	Kuroda *et al.*, 2001	5 (CC5)
Mu50 (hu2)	Human (surgical wound infection)	Japan	1997	Kuroda *et al.*, 2001	5 (CC5)
C126P (hu7)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	12 (CC12)
MRSA252 (hu9)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	36 (CC30)
H383 (hu8)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	22 (CC22)
C434 (hu6)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	8 (CC8)
MSSA476 (hu5)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	1 (CC1)
D117 (hu10)	Human (blood isolate)	England (UK)	1997	Enright *et al.*, 2000	30 (CC30)
MSA2020 (hu11)	Human (Scalded Skin Syndrome)	France	Unknown	Fitzgerald *et al.*, 2003	121 (CC121)
951 (bo1)	Bovine (mastitis)	USA	1990	Sischo *et al.*, 1993	126 (CC126)
C123/5/005 (bo2)	Bovine (milk isolate)	UK	2003	Sung *et al.*, 2008	151 (CC151)
V329 (bo3)	Bovine	Spain	Pre-2001	Cucarella *et al.*, 2001	133 (CC133)
RF103 (bo4)	Bovine (clinical mastitis)	Ireland	1993	Fitzgerald *et al.*, 1997	71 (CC97)

Strain *	**Host (disease or characteristic)**	**Geographic origin**	**Isolation date**	**Reference**	**MLST (clonal complex)**
891-1 (ov1)	Ovine (milk isolate)	Norway	1998-2004	Mork *et al.*, 2005	133 (CC133)
1 (ov2)	Ovine (milk isolate)	Denmark	1998-2004	Mork *et al.*, 2005	9 slv (CC9)
6659-2 (ov3)	Ovine (milk isolate)	Sweden	1998-2004	Mork *et al.*, 2005	151 slv (CC151)
ED133 (ov4) †	Ovine (mastitis)	France	1997	Ben Zakour *et al.*, 2008	133 (CC133)
VET-BZ30 (ov5)	Ovine (subclinical mastitis)	Brazil	1996/97	Aires-de-Sousa *et al.*, 2007	750 (CC133)
Newman	Human	UK	1950s	Duthie & Lorenz, 1952	ST8 (CC8)
Newman Δ*spA*	Manipulated strain deficient in SpA	N/A	N/A	Mempel *et al.*, 1998	ST8 (CC8)

Names in parentheses indicate host origin of strains selected for analysis of MGE distribution (Chapter 4)

† Previously published with the strain name 1174

described previously (Enright *et al.*, 2000), except for the *tpi* locus where alternative oligonucleotide primers were utilised (F, 5'-GCATTAGCAGATTTAGGCGTTA-3' and R, 5'-TGCACCTTCTAACAATTGTACGA-3') (Armand-Lefevre *et al.*, 2005). PCR purification, directed sequencing and *in silico* sequence analysis were performed as described in materials and methods. An allelic profile was obtained for each strain by comparing sequence for the seven gene fragments with the database on the *S. aureus* MLST website (http://saureus.mlst.net).

ST was determined by allelic profile, and each ST assigned to a CC following e-BURST analysis of all STs added to the online database (Feil *et al.*, 2004). Where an allele was found to be novel (i.e. not represented by any strains on the database), directed sequencing reads generated using both forward and reverse primers were submitted to database curators to obtain a unique allele number. Where an allelic profile was found to be novel, details were submitted to obtain a unique ST.

3.3.3. Phylogenetic analysis

Reconstruction of evolutionary relationships between avian isolates was performed as described in materials and methods. Analysis was performed on the concatenated sequences of internal fragments of seven housekeeping genes, as utilised in the MLST method. SNP discovery and SNP-based phylogenetic analysis of the ST5 radiation was performed by Ulrich Nubel and colleagues at the Robert Koch Institute, Germany, as previously described (Nubel *et al.*, 2008).

3.3.4. Extraction of cell wall-associated proteins

Cells from 1 ml volume of bacterial culture were harvested by centrifugation at 4000 rpm for 8 min and washed once with PBS. The pellet was re-suspended in lysis buffer (50 mM Tris/HCl, 20 mM $MgCl_2$, pH 7.5, supplemented with 30 % (w/v) raffinose) and cells lysed by incubation with lysostaphin (AMBI products LLC, 200 µg ml^{-1}) at 37 ºC for 20 min, with addition of protease inhibitors (Sigma Protease Inhibitor Cocktail, 250

µl g^{-1} cell wet weight). Protoplasts were pelleted by centrifugation at 6000 x *g* for 20 min, and supernatant containing solubilised cell wall-associated proteins removed.

3.3.5. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Successful extraction of protein from cells was confirmed by SDS-PAGE. Cell-wall associated protein samples were heated at 95 ºC for 5 min with equal volume of protein denaturing sample buffer (62.5 mM Tris-HCl p.H 6.8, 20 % (w/v) glycerol, 2 % (w/v) SDS, 5 % (w/v) mercapthoethanol). Samples were loaded into wells of 12 % (w/v) polyacrylamide gel, with protein marker (Full Range Rainbow Molecular Weight Marker, Amersham Biosciences) in adjacent wells. Electrophoresis was carried out at 50 mA with Tris-glycine running buffer (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.2) using a Hoefer SE600 PAGE unit, for 4 h. Gels were stained with Coomassie Blue stain and destained with H_2O to visualise protein bands.

3.3.6. Western immuno-blot analysis

Samples were separated by SDS-PAGE and electro-transferred to nitrocellulose membrane (Amersham Hybond ECL) in BioRad transfer unit, with transfer buffer (25 mM Tris-HCl pH 7.6, 192 mM glycine, 20 % (v/v) methanol) at 10 V overnight. The membrane was incubated for 1 h with blocking buffer (8 % (w/v) 'Marvel' milk powder in PBS). This was followed by 2 h incubation with affinity-purified chicken-anti-Protein A antibody (Gallus Immunotech) at 1:10,000 dilution, 3 x 20 min washes with washing buffer (0.05 % (v/v) Tween 20 and 1 % (w/v) 'Marvel' milk powder in PBS), 1 h incubation with horseradish peroxidase-conjugated affinity-purified goat anti-chicken IgY Fab fragment antibody at 1:5,000 dilution, and further 3 x 20 min washes with washing buffer. Immunoreactive proteins were identified by enhanced chemiluminescent (ECL) detection, with radiographic film (Amersham Biosciences Hyperfilm ECL) developed in an X-Ray film processor (Optimax).

3.3.7. PCR amplification, sequencing and bioinformatic analysis of the *spA* gene

A 2 kbp region comprising the *spA* gene and 600 bp of intergenic sequence was amplified by PCR (for primer sequences see Table 4), purified, resolved by gel electrophoresis, sequenced by directed sequencing and analysed *in silico*, as described in the materials and methods section. A unique pair of oligonucleotide primers was designed for amplification of the *spA* gene in AV86, due to sequence variation in the primer annealing site (Table 4). Sequence reads were assembled into contigs and aligned using Vector NTI software (Invitrogen), and alignments analysed to identify SNPs or variation in number of repeats.

3.4. Results

3.4.1. The majority of *S. aureus* isolates from poultry belong to clonal complex 5 (CC5)

In order to examine the population genetics of *S. aureus* strains infecting farmed broiler chickens, MLST was performed with 57 *S. aureus* isolates including 48 isolates from healthy and diseased poultry in eight countries on four continents isolated over the last 54 years (Table 3). Also included in the study were 9 isolates from different species of reared game and wild birds (Table 3). Remarkably, the majority of all avian isolates (n=35; 61 %) including 32 (67 %) from broiler chickens belonged to a single sequence type, ST5, or its single locus variants (ST1342, ST1346 and ST1350) including isolates from all countries examined except Australia (Fig. 7). CC5 is a well characterised lineage of *S. aureus*, responsible for disease in humans worldwide (Nubel *et al.*, 2008).

Of the non-CC5 isolates identified, which included isolates from broiler chickens, reared bird species such as pheasant and partridge, and a wild buzzard, 11 (19 %) belonged to

Table 4. Oligonucleotides designed for PCR amplification and sequence analysis of *spA* gene

Name*	Sequence (5'-3')	Reactions requiring primer
spa-f	TCTCTATTACGCAAGTGTGC	PCR and sequencing
spa-f2	CGAAATAGCGTGATTTTGCGG	Sequencing
spa-f3	CCCAAGCCAAAGCACTAATG	Sequencing
spa-f4	ATCTGGTGGCGTAACACCTG	Sequencing
spa-r	ACAAAAGATGTTGCTCGTGC	PCR and sequencing
spa-r2	CATTAGTGCTTTGGCTTGGG	Sequencing
spa-r3	CATGTACTCCGTTGCCGTCT	Sequencing
spa-r4-b	AGCAGTAGTGCCGTTTGCTT	Sequencing
spa-av86-f	TTAAATATGCCTAACTTAAATGCTGA	Sequencing
spa-av86-r	AGCAGTAGTGCCGTTTGCTT	Sequencing

* f, forward primer; r, reverse primer

CC385. Phylogenetic analysis of concatenated MLST loci, from all avian isolates included in the study and representatives of prominent human and other animal lineages, demonstrates that CC385 is not closely related to CC5 (Fig. 8). Furthermore, CC385 strains have not been previously isolated from human or other mammalian hosts.

Of the remaining 11 isolates, 4 from Australia belonged to a novel ST1345, 3 Japanese isolates were ST1027, 2 were ST8, and single isolates were ST1 and ST15. Scrutiny of the MLST online database reveals that these lineages all contain strains of human origin, or are close relatives of human associated lineages.

Seven novel MLST sequence types were identified in this study including ST1342, ST1343, ST1344 and ST1345 (poultry isolates), and ST1346, ST1347 and ST1350 (other avian isolates). Comparison with the MLST database revealed that ST1342, ST1346 and ST1350 are single locus variants of ST5, and that ST1343, ST1344 and ST1347 belong to CC385 (single, double or triple locus variants).

3.4.2. Human and poultry isolates of the CC5 lineage have different phenotypes

A key phenotypic characteristic distinguishing poultry isolates from human isolates is that poultry strains do not produce the surface protein SpA (Isigidi *et al.*, 1990). In order to investigate the production of SpA by poultry isolates, cell wall-associated proteins were extracted from a subset of the poultry isolates (n=16), including 14 of the CC5 lineage and 2 of CC385, at mid-exponential and at stationary phase of growth. Separation of cell wall associated proteins by SDS-PAGE was followed by Western blot analysis with SpA-specific IgG. In addition, proteins were extracted from a human CC5 strain (N315), a human CC8 isolate (Newman) and an isogenic mutant derivative strain deficient in SpA (Newman Δ*spA*).

Poultry isolates of the CC5 lineage did not express detectable SpA at exponential or stationary phases, in contrast to CC5 human isolate N315 which expressed the protein at

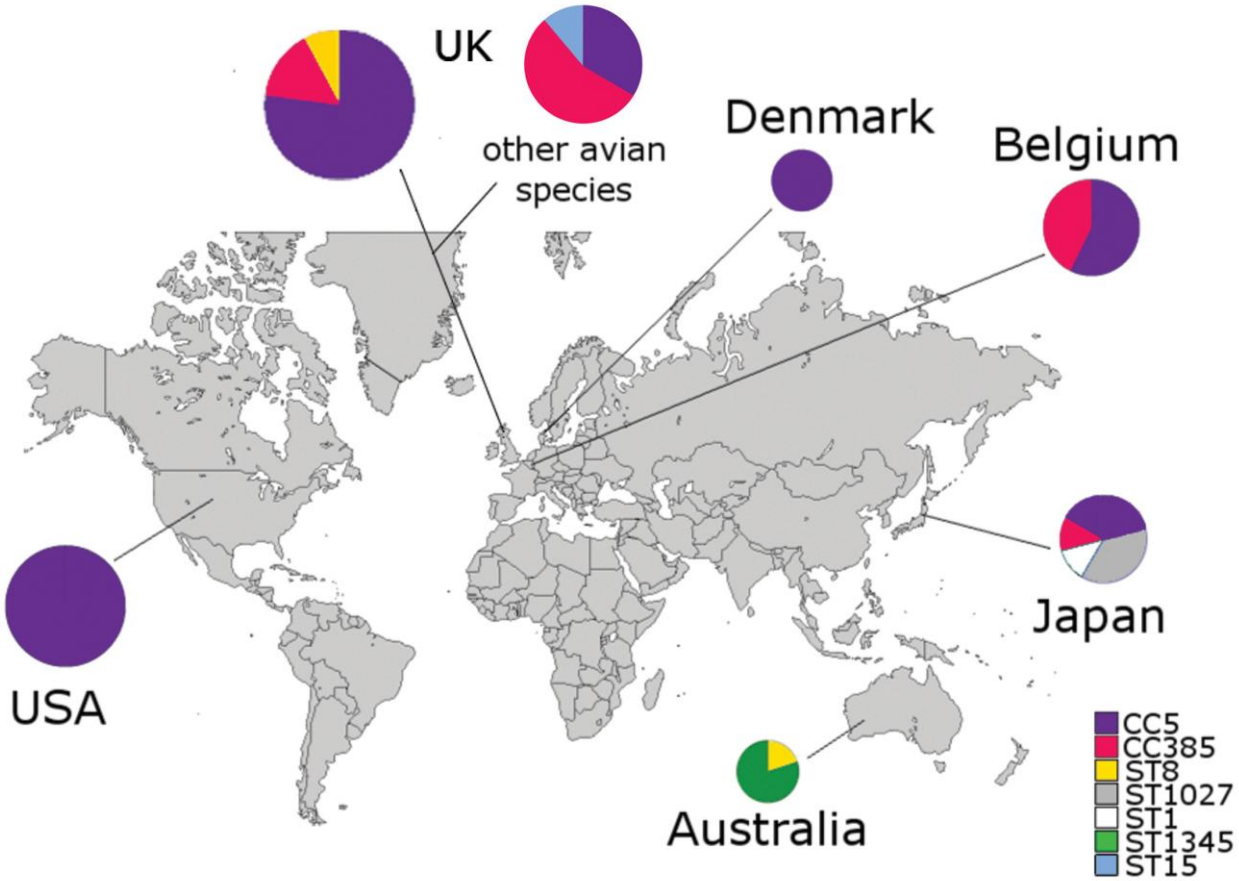


Figure 7. The majority of poultry *S. aureus* isolates belong to the single, human associated lineage CC5. Geographic distribution of poultry and other avian *S. aureus* Sequence Types (STs) identified in countries on 4 different continents. Pie chart diameter is proportional to the number of isolates from each location.

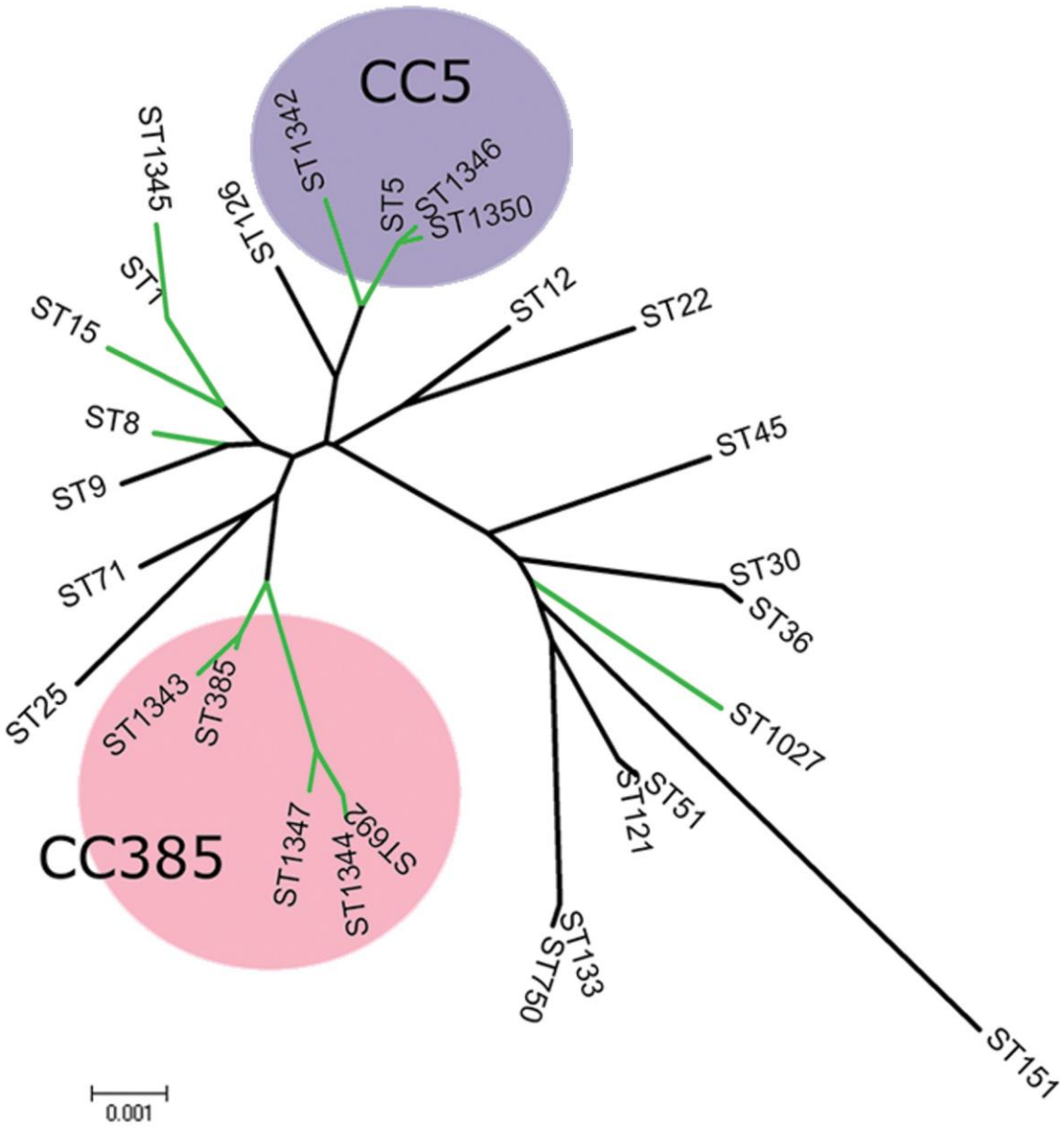


Figure 8. The majority of poultry isolates fall into two unrelated lineages, CC5 and CC385. Neighbour joining tree with bootstrapping consensus inferred from 500 replicates, constructed using concatenated sequences of *S. aureus* STs from birds (green branches), and representatives of the major clonal complexes of human and animal origin in the *S. aureus* species (black branches). The 2 major avian-associated clonal complexes, CC5 and CC385, are indicated in purple and pink, respectively.

both phases (Fig. 9). Poultry isolates of the CC385 lineage express SpA at stationary phase but not at mid-exponential (Fig. 9).

To ascertain the genetic basis for the observed variation in expression of a major surface protein in the CC5 lineage, the *spA* gene that encodes the protein was amplified by PCR and sequenced using a primer walking approach. A SNP at position 655 ($G\text{-}T^{655}$) was identified in all ST5 strains (Fig. 9) resulting in a premature stop codon. The mutation occurs only 219 amino acids from the start codon, suggesting that no functional protein is produced.

3.4.3. The poultry ST5 clade is the result of a single human to poultry host jump.

In a previous study, Nubel *et al.* (2008) performed a high resolution analysis of the phylogenetic structure of the human CC5 clonal radiation, by mutation discovery at 108 loci (46 kbp). This resulted in the identification of at least 14 distinct lineages within the CC5 group (Nubel *et al.*, 2008). Investigation of phenotypic differences between poultry and human strains of the CC5 clade suggests that the two groups of isolates may belong to distinct subgroups.

Mutation discovery was performed at the same genetic loci among a representative selection of 19 poultry CC5 isolates and 29 novel bi-allelic polymorphisms (BiPs) were identified. These polymorphisms, along with 156 informative BiPs previously identified among human CC5 isolates (Nubel *et al.*, 2008), were used to construct a minimum spanning tree which included the 19 poultry isolates and 135 isolates of human origin (Fig. 10). The phylogeny suggests that the poultry CC5 clade belongs to a single sub-lineage (Fig. 10).

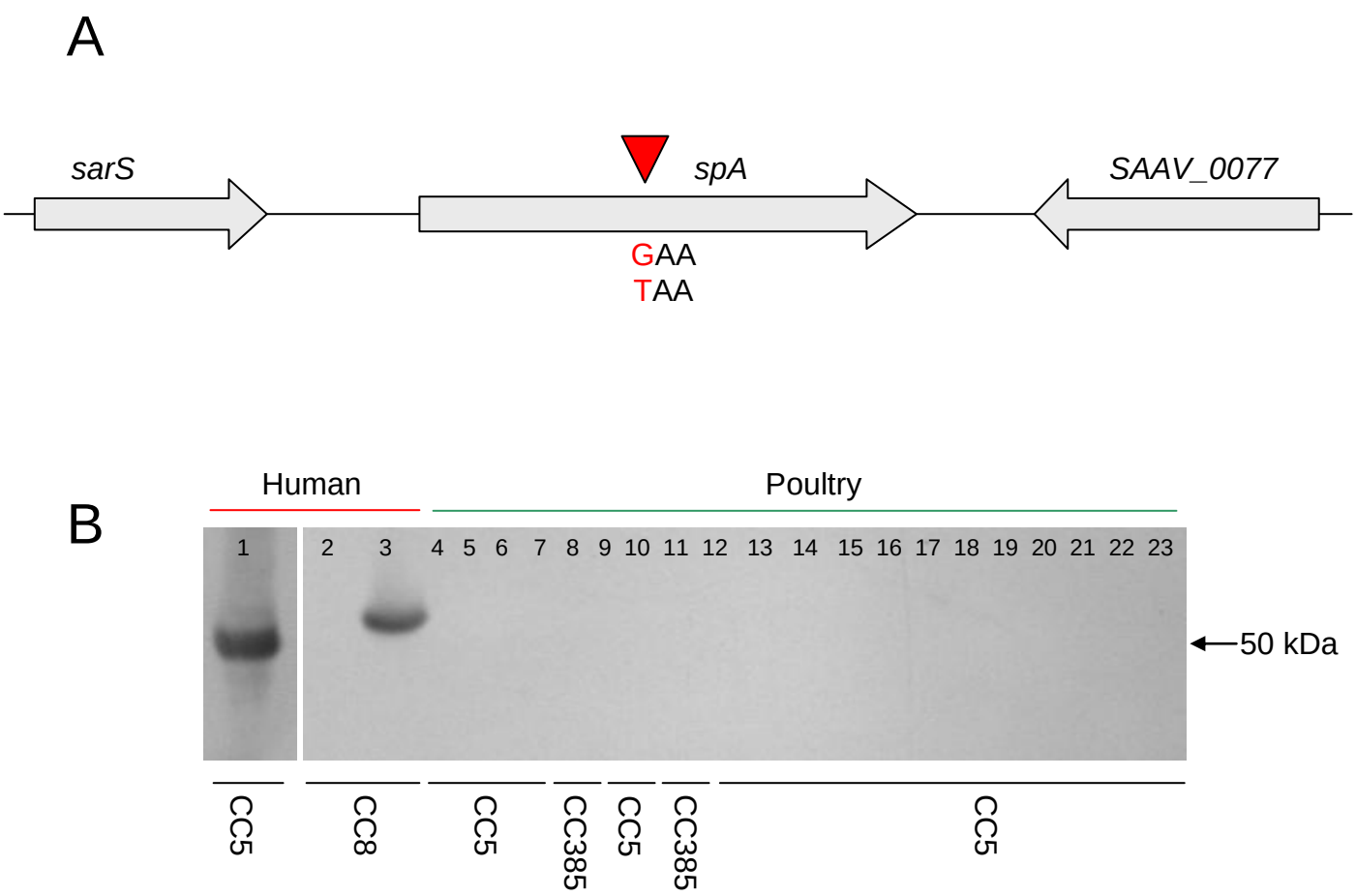


Figure 9. Poultry *S. aureus* isolates of the CC5 lineage do not express the surface protein Staphylococcal Protein A (SpA). (A) Amplification by PCR and directed sequencing of the spA gene reveal a mutation at position 655 (G-T^{655}) in CC5 poultry strains, resulting in a premature stop codon. **(B)** Western immuno-blot of cell-wall associated proteins using antibodies against SpA. Proteins extracted at mid-exponential growth phase. 1, N315; 2, NewmanΔ*spA*; 3, Newman; 4, AV4; 5, AV18; 6, AV24; 7, AV35; 8, AV86; 9, AV72; 10, CIX2; 11, CIX18; 12, CIX38; 13, CIX51; 14, 10009; 15, 007439; 16, 008078; 17, 008661; 18, 009705; 19, 997184; 20, 335; 21, 495; 22, 499; 23, 634.

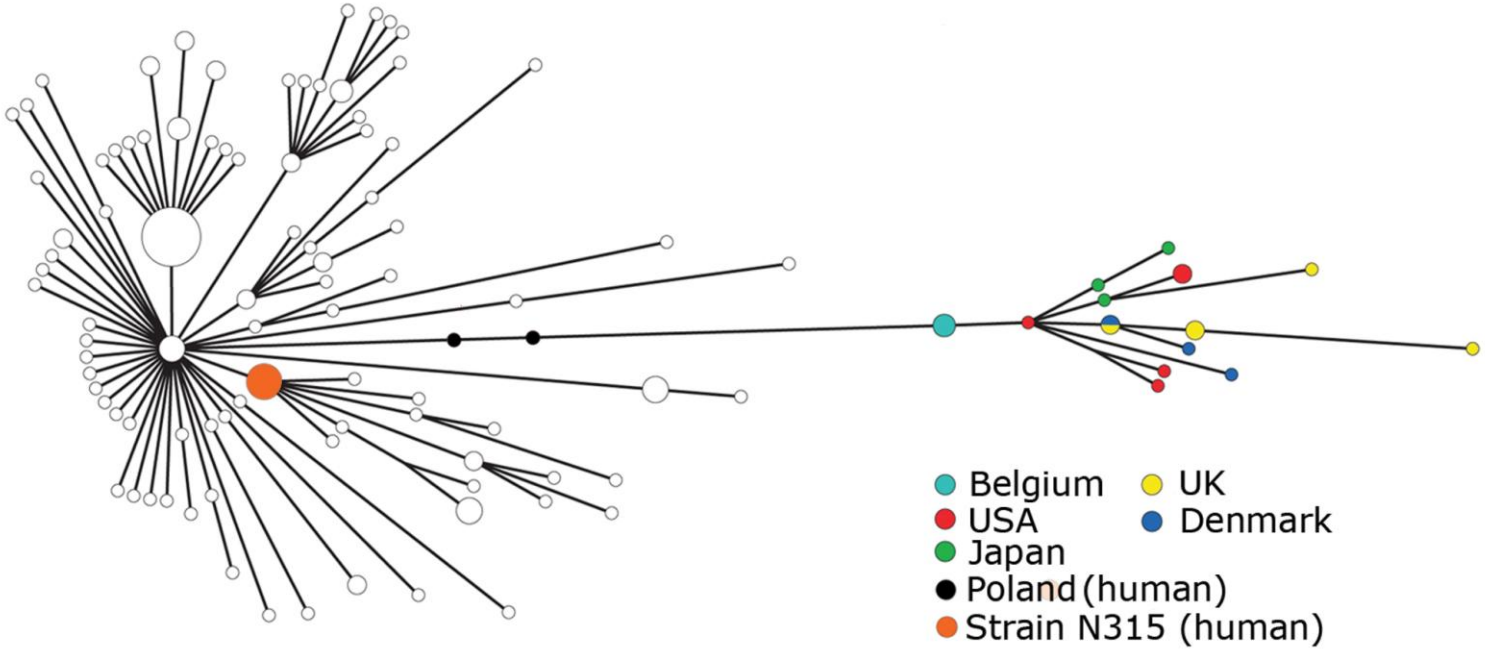


Figure 10. The majority of poultry *S. aureus* isolates evolved by a single host jump from humans followed by wide dissemination. Minimum spanning tree based on 185 bi-allelic polymorphisms in 19 representative poultry ST5 isolates compared to 135 *S. aureus* ST5 human isolates of global origin analyzed in a previous study (Nubel *et al.*, 2008). Circle size is proportional to haplotype frequency and line length is proportional to the number of mutational steps between haplotypes. Coloured nodes indicate the geographical origin of the poultry isolates and closely-related human isolates from Poland (strains PL72 & MR1), and the node representing the human strain N315. For reference, the distance between the two polish human isolates (black circles) is one SNP. Belgian strains (light blue circle) and strain ED98 (yellow circle to far right of figure) differ by six SNPs.

3.4.4. The poultry CC5 clade has undergone rapid intercontinental dissemination.

Following phylogenetic analysis of human CC5 isolates, Nubel *et al.* (2008) observed strong geographical segregation whereby sub-lineages were generally restricted to one area. The distribution of the poultry ST5 clade across countries on 3 different continents strongly contrasts with the isolation of the basal human ST5 isolates to Poland (Fig. 10), indicating frequent intercontinental spread of the poultry strains.

3.4.5. Commensal isolates and invasive disease isolates co-segregate

The collection of *S. aureus* strains examined included both commensal isolates (22 % of ST5 strains) and those from invasive disease (75 %), but these did not segregate following SNP based population genetic analysis (Fig. 10) suggesting that broiler poultry infections are caused by resident commensal strains of *S. aureus*.

3.5. Discussion

The majority of avian isolates (n= 35; 61 %) including 32 (67 %) from broiler chickens belonged to CC5 (Fig. 7). This is one of the most successful human-associated lineages of *S. aureus*, characterised by its global distribution and the frequent emergence of methicillin-resistant strains (Nubel *et al.*, 2008). In this respect, poultry isolates contrast sharply with those from a bovine or ovine host which typically belong to host-specific lineages (Smyth *et al.*, 2009). This finding suggests that the CC5 lineage has switched host relatively recently, either from humans to poultry or in the opposite direction.

Smyth *et al.* (2009) also made this observation following typing of a panel of animal isolates. They found that isolates from both poultry and rabbit hosts belonged to lineages normally associated with humans, suggesting the possibility of cross-species

transfer. However, their study considered only 12 poultry strains, all of which were from Northern Ireland.

Phenotypic variation between human and poultry strains regarding SpA expression has been acknowledged for over 20 years, (Isigidi *et al.*, 1990). Results of the current study indicate that strains of the CC5 poultry clade do not express SpA due to a single mutation that renders the *spA* gene a pseudogene. Of note, avian strains of the CC385 lineage have an intact gene and express this protein in stationary phase, though concentration of SpA is markedly reduced compared to human CC5 and CC8 strains. The reduced expression level of SpA in CC385 may suggest a regulatory adaptation.

SpA has multiple roles in the pathogenesis of human infections, including adherence to lung epithelium via the TNFR1 receptor (Gomez *et al.*, 2006), binding to human platelets through its interaction with von Willebrand factor (Hartleib *et al.*, 2000, O'Seaghdha *et al.*, 2006), and inhibition of opsonophagocytosis resulting from non-specific binding of the Fc region of IgG (Foster, 2005, Gemmell *et al.*, 1991). Of note, the avian equivalent of human IgG is IgY, which has a structurally distinct Fc region that does not bind to SpA (Barkas & Watson, 1979, Warr *et al.*, 1995), suggesting that SpA would not contribute to the inhibition of opsonophagocytosis in birds. It is possible, therefore, that loss of the *spA* gene could be an adaptation to an avian host.

In order to trace the evolutionary origin of the poultry-associated CC5 clade identified in this study, bi-allelic polymorphisms discovered in both human and poultry isolates were used to construct a minimum spanning tree. This included 19 poultry isolates and 135 isolates of human origin (Fig. 10). The tree indicates that the poultry CC5 clade belongs to a CC5 sub-lineage which includes human strains that were circulating in Poland in the 1990s (Fig. 10). Importantly, examination of the distribution of SNPs within the lineage indicates that the Polish human isolates are basal to all isolates in the poultry ST5 clade (Fig. 10).

The great diversity of geographically isolated subtypes of the CC5 lineage colonising humans indicates a long term association of the CC5 lineage with its human host (Nubel *et al.*, 2008) and strongly suggests that the common poultry *S. aureus* clade evolved as a result of a human-to-poultry host jump by a *S. aureus* strain originating in or near Poland. This led to the emergence of a new strain with a tropism for birds. The alternative scenario of a poultry-to-human host jump could only be explained by multiple host species jumps including an initial jump from humans into an unknown non-human intermediate host species which occurred since the most recent common ancestor of the CC5 lineage. This would be followed by a second jump into poultry, resulting in the poultry ST5 clade. A final jump back into humans would be required leading to wide dissemination in Polish hospitals (Leski *et al.*, 1998).

In their study of human isolates, Nubel *et al.* (2008) previously found that CC5 sub-lineages exhibit phylogeographic clustering, including many which are confined to single countries. This indicates that intercontinental spread and fixation in local populations by human *S. aureus* is very rare. For example, the widespread clone represented by Polish human isolates MR1 and PL72, which was predominant in Polish hospitals in the 1990s (Leski *et al.*, 1998), has not been identified anywhere else in the world. Geographical restriction also predominates in the ST239 clonal lineage, a clade that accounts for a large proportion of healthcare-associated human infections throughout Asia and Eastern Europe (Harris *et al.*, 2010). Strains with the same geographic origin largely clustered together in a SNP-based phylogenetic tree, with separate clades for South America, Asia (Thailand and China) and Greece, among others (Harris *et al.*, 2010).

The distribution of the poultry CC5 clade in countries on 3 different continents strongly contrasts with this (Fig. 10), indicating that frequent intercontinental spread has been a prevalent process, shaping the current population of chicken-associated *S. aureus*. This finding is consistent with the globalised nature of the poultry industry, whereby a very small number of breeder companies distribute large numbers of live broiler chickens via

a global distribution network. Interestingly, CC5 isolates were identified in all countries examined except for Australia (Fig. 7). This could be attributed to the country's stringent restrictions regarding the import and export of livestock. Until the 1990s poultry imports were banned entirely, and quarantine restrictions are now very restrictive (see http://www.daff.gov.au/).

Of the non-CC5 isolates identified, which included isolates from broiler chickens as well as reared and wild bird species, 11 (19 %) belonged to CC385 (Fig. 7). This lineage is not related to the CC5 group (Fig. 8) and has not been previously identified among human or mammalian strains. The lack of association of CC385 with other host species and its wide distribution among an array of avian species in different countries (Fig. 8, Table 3) suggests that this lineage has a long-term avian host association, and may possess traits advantageous for infection of this host taxon**.**

Of the remaining 11 isolates, 4 from Australia belonged to a novel ST1345, 3 Japanese isolates were ST1027, 2 were of ST8, and single isolates were of ST1 and ST15 (Fig. 7). All of these STs are related or identical to those previously identified among human *S. aureus* isolates. It is possible, therefore, that they represent other independent occurrences of a human to poultry host switch or vice versa.

The collection of *S. aureus* strains examined included both commensal isolates and those from invasive disease, but phylogenetic analysis did not distinguish between the two groups (Fig. 10). This provides support for the hypothesis that *S. aureus* is an opportunistic pathogen which colonises the skin, nasal epithelium or equivalent site, but has the potential to invade deeper tissue and cause disease, consistent with the findings of previous studies of human isolates (Feil *et al.*, 2003, Lindsay *et al.*, 2006). In humans the opportunity for bacterial invasion can arise due to invasive medical procedures such as catheter insertion, surgery, or a weakened immune system due to other conditions (Anderson & Kaye, 2009, Davidson & Boucher, 2009, Yanagi & Sheretz, 2009). In poultry, intensive farming practices often include very high population densities, limited

environmental stimuli and a lack of individual attention to identify injury. It is possible that such conditions increase the likelihood of the epidermis being breached. Furthermore, any host predisposition for susceptibility to invasion will be present in most individuals of the genetically homogeneous broiler chicken population, which has been subject to selective breeding over many years. Alternatively, the CC5 poultry clade may possess genes that confer a phenotype particularly adept in invasion of a poultry host, and of the leg tissue in particular. Of note, a previous study reported that the ST5 clonal complex is associated with increased frequency of haematogenous infections of humans, including osteomyelitis (Fowler *et al.*, 2007), consistent with the largely skeletal tropism of infections caused by the poultry ST5 clade. Such factors may explain the observed increase in prevalence of BCO, a condition that has become a leading cause of lameness in the broiler chicken industry (McNamee & Smyth, 2000).

In conclusion, this study reveals that the majority of *S. aureus* isolates from broiler chickens are the descendants of a single human-to-poultry host jump by a subtype of the worldwide human CC5 clonal lineage unique to Poland. In contrast to human subtypes of the CC5 radiation, which demonstrate strong geographic clustering, the poultry CC5 clade was distributed in different continents consistent with wide dissemination via the global poultry industry distribution network.

4. Comparative genomic analysis reveals avian host adaptation by *S. aureus*

4.1. Introduction

Phylogenetic analysis of a collection of poultry isolates from 6 countries on 4 continents revealed that the majority are part of the CC5 lineage usually associated with humans (Chapter 3). They form a poultry-specific clade, though two Polish human associated MRSA strains fall into the same sublineage, with phylogenetic analysis suggesting that they are basal to the poultry group. These data suggest that the poultry clade is the result of a single human-to-poultry host jump. Further analysis of genotypic and phenotypic variation between the two host associated clades of the CC5 group is required in order to elucidate the molecular basis for the poultry host tropism.

Genome sequencing and comparative analysis have had a profound impact on the field of microbiology, revealing a great deal about the evolutionary history of microbial pathogens. The aim of this investigation was to utilise next generation sequencing techniques to determine whole genome sequences for two CC5 *S. aureus* isolates, poultry isolate ED98 and Polish human isolate MR1.

Direct comparison of the two strains was performed in order to identify changes associated with adaptation to a novel poultry niche. Furthermore, comparison with sequences deposited in the NCBI Genbank database, including a number of *S. aureus* genomes, facilitated identification of unique genomic regions or alleles resulting from genetic drift or natural selection.

4.2. Aims

To determine the underlying genomic basis for adaptation to an avian host by the CC5 poultry clade, following a human-to-poultry host jump:

- To sequence the genome of *S. aureus* strain ED98, a representative of the CC5 poultry clade.
- To sequence the genome of *S. aureus* strain MR1, a human MRSA isolate basal to the CC5 poultry clade.
- To perform comparative genomic analysis of ED98 with MR1, and with other published *S. aureus* sequences, in order to identify novel regions and/or alleles associated with a particular host.

4.3. Materials and methods

4.3.1. Genome sequencing

Massively parallel 454 pyrosequencing of poultry strain ED98 was carried out by 454 LifeSciences (www.454.com) with a GS20 sequencer followed by initial assembly using the Newbler program (Roche, Sussex, UK). The order and orientation of assembled contigs was determined by scaffolding against the published genome sequence of the human ST5 strain N315 (GenBank BA000018). Primers specific for sequence gap edges were designed using Projector 2 (van Hijum *et al.*, 2005) and used for PCR amplification with Platinum PCR Supermix (Invitrogen, Paisley, UK) or PFU Ultra II Fusion HS DNA Polymerase (Promega, Hampshire, UK) in a Biometra TGradient, followed by directed sequencing of PCR products by primer walking. Regions containing duplicated sequences or internal repeats were amplified by PCR and cloned into the pSC-A-amp/kan or pSC-B-amp/kan vectors using the StrataClone PCR cloning kit or Blunt PCR cloning kit, respectively (Agilent Technologies, West Lothian, UK). Directed sequencing was then performed using specific primers and a plasmid DNA template, facilitating longer sequencing reads for spanning repeat regions.

Sequence reads and flanking pair of contigs were first assembled individually using Vector NTI contig express software (Invitrogen) or the Phred/Phrap/Consed package (Ewing *et al.*, 1998, Gordon, 2004), to ensure that assembly produced a fragment of the same size as the amplified PCR product. Finally, whole genome assembly of contigs

(generated from shotgun sequence reads) and directed sequencing reads was performed using the Phred/Phrap/Consed package (Ewing *et al.*, 1998, Gordon, 2004).

Genome sequencing of *S. aureus* strain MR1 was performed with an Illumina Solexa Genome Analyzer (www.illumina.com) at The Gene Pool sequencing facility (University of Edinburgh). Raw sequence reads were aligned against the ED98 genome using MAQ (http://maq.sourceforge.net) to identify SNPs. Sequences failing to map to the ED98 reference were assembled into contigs using Velvet (Zerbino & Birney, 2008). For the purposes of submission to the NCBI database, reads were also assembled into contigs *de novo* in a separate procedure. This was performed using Velvet (Zerbino & Birney, 2008).

4.3.2. Genome annotation and comparative analysis

Where nucleotide sequence of poultry strain ED98 exhibited >85 % homology to human strain COL, gene annotation was duplicated from the published annotation of COL using GATU software (Tcherepanov *et al.*, 2006). The remaining genes were annotated using the J. Craig Venter Institute (JCVI; http://www.jcvi.org) automated pipeline annotation which includes Glimmer, BLAST Extend Repraze (BER) alignments, alignments with experimentally characterized genes, HMM matching, and searches for biologically significant patterns with PROSITE. Data were then manually curated in Manatee (http://manatee.sourceforge.net) and Artemis (Rutherford *et al.*, 2000). Complementary annotation data were provided by the SEED (Overbeek *et al.*, 2005) and the RAST annotation server (Aziz *et al.*, 2008), and start sites were curated for all genes. Further analysis of several individual genes of interest was performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and interpro scan (Zdobnov & Apweiler, 2001). Pairwise comparisons between ED98 and published genomes were performed using the Artemis Comparison Tool (Carver *et al.*, 2005).

Gene content of strain MR1 was examined using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), dot matrix alignments and the Artemis

Comparison Tool (Carver *et al.*, 2005). Prophages were classified as described previously based on conservation of nucleotide sequences across entire phage (Kwan *et al.*, 2005). Comparative analysis between ED98 and MR1 at the level of single nucleotides was performed using MAQ software. SNPs identified were filtered to exclude those in regions containing MGE (as they are unlikely to be orthologous) and those with lower than average depth of coverage (<90 sequence reads) or mapping value.

4.3.3. Pseudogene identification

ORFs displaying evidence of frameshifts were indicated by the BER search performed in the JCVI annotation pipeline and manually checked. In addition, a Perl script was written for identification of truncated ORFs in ED98 caused by mutations leading to a premature stop codon or indels causing a frameshift. The sequence of each gene of ED98 was compared to its closest homolog as determined by the BER search. Size difference as a percentage of the gene length and percentage identity to best hit was reported and regions manually checked to eliminate differences caused by variable start site predictions. Script preparation was performed by Nouri Ben Zakour (University of Edinburgh). Mutations were confirmed by directed sequencing.

The distribution of pseudogenes was examined in 19 additional poultry CC5 isolates and the human strains MR1 and N315, the latter of which had been analysed for the presence of pseudogenes in a previous study (Lerat & Ochman, 2005). Primers were designed for amplification of the region surrounding each mutation (Table 5). Reactions were performed as described in materials and methods using *GoTaq* DNA polymerase (Invitrogen, Paisley, UK), with an initial 2 min denaturation at 95 °C, followed by 30 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min with a final incubation at 72 °C for 5 min.

4.3.4. PCR detection of MGE

The presence of novel MGE among a panel of avian, human, bovine and ovine isolates

Table 5. Oligonucleotide primers used for PCR detection of avian *S. aureus* mobile genetic elements (MGE) and sequencing of mutations that lead to pseudogene formation in ED98.

Target		Forward primer	Sequence 5'-3'	Reverse primer	Sequence 3'-5'	Product size
MGE	φAv1	φAv1-F3	TACAACGTCCGATTCAGCAA	φAv1-R2	GAGCGAAATTTCTGGGACTG	3180bp
	φAvβ	φAvβ -F1	GCTTTGACATTTTCGGCATT	φAvβ -R1	CTGGCTTGCTTTCTTCTGCT	1646bp
	φAvβ	φAvβ -F2	TTAACCCACTCCGCAAATTC	φAvβ -R2	ATCGCAAGAGACCAAGCAGT	3754bp
	SaPIAv	SaPIAv-F2	TGAATGGGCAACAGACGATA	SaPIAv-R	TGATCAGCGTAAGTGGATGC	3225bp
	SaPIAv	SaPIAv-F1	TCGCTGATTTCTTTGTCACG	SaPIAv-R	TGATCAGCGTAAGTGGATGC	1470bp
	pAvY	pAvY-Lt	CATTGAGCCGAATTTGATTG	pAvY-Rt	TTTAACTGCCCATTTTCATCG	287bp
	pAvX	Thiol-F	TTTTCTGGTCGCCATACTCC	Thiol-R	TCCATCGGCAGTTAAAAAGC	1148bp
Pseudogenes	*spA*	Spa-f	TCTCTATTACGCAAGTGTGC	Spa-r2	CATTAGTGCTTTGGCTTGGG	732bp
	asp1	Asp1-f	ATCCACAAGTCGTTGCATCA	Asp1-r	CAGGGAGAGGCTTCTTACCA	545bp
	sspA	Sspa-f	CCCCTAATGAGCAAAACAAA	Sspa-r	TCATTAATAAATACCGCACCATT	299bp
	SAAV_1787	SAAV1787-f	TCCATTGTTACCTCCCGATT	SAAV1787-r	TGGTTGTTCAGTTAATGTTGGTG	479bp
	SAAV_0389	SAAV0389-f	AGCGAAGGTGGATGAAGATG	SAAV0389-r	TCCCAGTCATCCCAAGAGTT	289bp
	SAAV_1594	SAAV1594-f	GGAGAGTGGTGCATGACAAA	SAAV1594-r	CGCATGTGCAATCTCTTGTT	528bp
	SAAV_0046	SAAV0046-f	ACAGGGAAAGAAGCGGAAAT	SAAV0046-r	TTATCGTGAGTGCGGAAGC	610bp

(Table 3) was determined by PCR using MGE-specific primers (Table 5). PCRs were performed as described in materials and methods using *GoTaq* DNA polymerase (Invitrogen, Paisley, UK). Reactions were performed with an initial 2 min denaturation at 95°C, followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min or 4 min (depending on predicted product size with each primer pair), with a final incubation at 72°C for 5 min.

4.4. Results

4.4.1. Genome sequencing of poultry and human isolates of the same CC5 sub-lineage

In order to determine the genomic basis of host adaptation following a human-to-poultry host jump, a complete genome sequence was obtained for poultry isolate ED98. This CC5 strain was isolated in 1996/97 from a chicken with BCO in Ireland (Table 3; Rodgers *et al.*, 1999). Shotgun sequencing reads were assembled into 53 contigs, of which 45 were mapped based on homology with human ST5 isolate N315, and 8 were unmapped. Closure of the gaps by PCR and directed sequencing resulted in a circular chromosome of 2,824,404bp (Genbank acc. no. CP001781) and three plasmids; pAvX at 17,256 bp (CP001782), pAvY at 1,442 bp (CP001783) and pT181 at 4,440 bp (CP001784) (Table 6).

Shotgun sequencing and *de novo* assembly into contigs was performed for human basal strain MR1, a hospital-acquired MRSA isolate from Poland (Leski *et al.*, 1998, Trzciński *et al.*, 1997). Contigs were deposited in the NCBI database for whole genome shotgun sequencing projects, under the accession number ACZQ00000000. This comprised one chromosome of approximately 2.8 mbp and plasmid pT181 at 4,440 bp, assembled into 339 contigs ranging in size from 200 bp to 108,324 bp (Table 6).

Table 6. **Comparative genome analysis of poultry isolate ED98 and basal human isolate MR1.**

Strain	Host	Size	GC %	Q40+	Novel MGE
ED98	Poultry	**Chromosome** 2,824,40 4bp **pAvX** 17,256 bp **pAvY** 1,442 bp **pT181** 4,440 bp	32.5 %	99.96%	φAvβ φAv1 SaPIAv pAvX pAvY
MR1	Human	**Chromosome** 2.8 mbp **pT181** 4,440 bp	32.7 %	n/d	SCCmec Type IV Class II clade C type phage Class II clade B type phage β-converting phage SaPImr1

4.4.2. The poultry CC5 clade has acquired novel mobile genetic elements from an avian-specific accessory gene pool.

In order to investigate the genetic basis for adaptation of the CC5 poultry clade to its new host, the whole genome sequence of poultry isolate ED98 was compared to the genomes of other *S. aureus* strains including basal human strain MR1 (Fig. 11; Table 7). This revealed that the poultry strain ED98 has acquired 5 novel MGE including 2 prophages, 2 plasmids and a SaPI (Fig. 11; Table 7). These MGE were not identified among human *S. aureus* sequences identified to date including the Polish MRSA strain MR1.

In addition, three copies of a transposon were identified in the ED98 genome, homologous (approximately 80 % nucleotide identity) to a transposon present in human associated strains MRSA252 (ST36), COL (ST250) and USA300 (ST8). This was absent from CC5 human isolates N315 (which is located within a different sub-lineage of the CC5 radiation) and MR1, indicating relatively recent acquisition or loss from strain MR1.

A panel of 48 *S. aureus* isolates from poultry, reared and wild birds, humans, cows, and sheep representing the breadth of diversity within the species was screened by PCR for the presence of the MGE identified in strain ED98. Results revealed the wide distribution of the novel MGE among avian *S. aureus* isolates and their complete absence among isolates of human or mammalian animal origin (Fig. 12).

Prophage φAvβ

A novel member of the β-converting phage family was identified, denoted φAvβ. Notably, this phage lacks the Immune Evasion Cluster (IEC) of genes which are central to the human niche-specific activity of most β-converting phages (van Wamel *et al.*, 2006). Instead of the IEC, φAvβ contains genes encoding a novel ornithine cyclodeaminase.

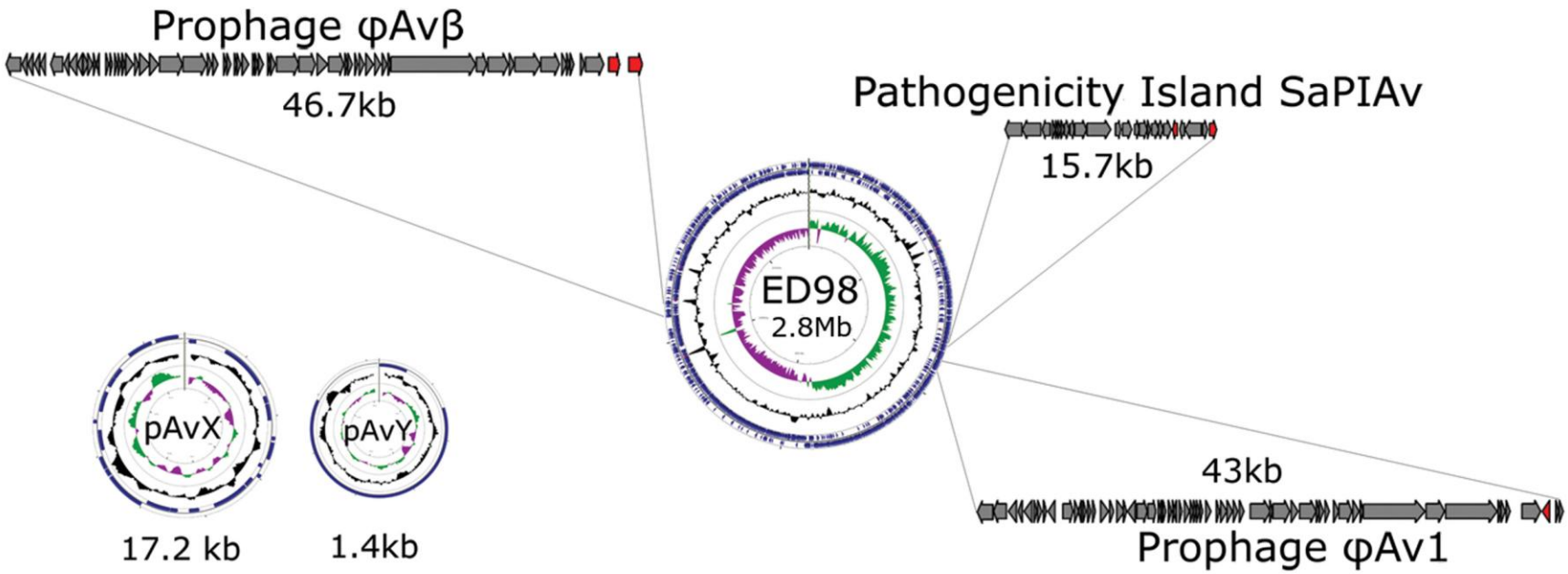


Figure 11. Novel MGE identified in the whole genome sequence of the ST5 poultry isolate ED98. The ED98 genome and plasmid maps represent ORFs (blue arrows on outer circle), GC content (middle circle) and GC skew (inner circle; GC+ and GC- skew in green and purple, respectively). Prophage and SaPI ORFs are represented by arrows, and red arrows denote genes putatively involved in virulence and/or host-specificity.

Table 7. Characteristics of the novel MGE identified in the genome sequence of the poultry strain ED98

MGE	Coordinates	Integration site	Similarity to other MGE	Genes putatively involved with virulence/ resistance / host specificity
φAvβ	2,041,200 - 2,087,967 bp	*hlb*	Integrase Sa3 type as found in other β-converting phage (Goerke *et al.*, 2009). Central region highly homologous to β-converting phage in MR1 (>90 % nucleotide identity across region of approximately 24 kbp). No or very low homology to other β-converting phage φN315, φNM3 or φSa3JH1 (>90 % nucleotide identity over region of approximately 3 kbp with φN315, but no homology over remaining approximately 42 kbp).	**Putative ornithine cyclodeaminase**: • 38 % amino acid identity to ornithine cyclodeaminase made by *Bacillus cereus*. • HMM match to ornithine cyclodeaminase/mu-crystalin family (PF02423). **Putative membrane protease**: • 27 % amino acid identity to Plnl (membrane-bound protease of CAAX family) made by *Lactobacillus plantarum.* • HMM match to CAAX amino terminal protease family.

MGE	Coordinates	Integration site	Similarity to other MGE	Genes putatively involved with virulence/ resistance / host specificity
φAv1	853,678 - 896,610 bp	Intergenic between *sufB* and SAAV_0879.	Large region of homology with 77-like phage (>90 % nucleotide identity across region of 20 kbp with φPVL). Region encoding Ear-like protein, holin and amidase is homologous (>95 % nucleotide identity) to region in other phage including φ71, and prophages in *S. aureus* strains JH1 and JH9.	**Ear-like protein**: • *ear* previously identified in pathogenicity islands SaPI1, SaPI3, SaPI5 and SaPImw2 (Subedi *et al.*, 2007). • *ear* encodes β-lactamase-like protein (Yarwood *et al.*, 2002).
SaPIAv	830,567 - 846,109 bp	Intergenic between SAAV_0786 and SAAV_0811.	Integrase and insertion site conserved with those of SaPI3 in *S. aureus* COL. Core region of 11 ORFs homologous (>90 % nucleotide identity) to characterized ORFs in SaPIbov. Two regions of novel hypothetical proteins at ends of island, as observed in other SaPI (Subedi *et al.*, 2007).	**Putative virulence region:** • Novel hypothetical proteins in accessory region A3 where virulence genes such as *tst* and *eta* located in other SaPI (Subedi *et al.*, 2007). • SAAV_0806: signal peptide, 1 transmembrane helix. • SAAV_0810: signal peptide, 4 transmembrane helices. • May suggest role as membrane transporters.

MGE	Coordinates	Integration site	Similarity to other MGE	Genes putatively involved with virulence/ resistance / host specificity
Transposon (3 copies with approximately 80 % identity)	52,827bp - 68,156bp 1,746,178 – 1,761,248bp 1,964,665 – 1,978,018,bp	SAAV_0046. Intergenic. Intergenic.	Homologous (approximately 80 % nucleotide identity) to transposon in human associated strains MRSA252 (ST36), COL (ST250) and USA300 (ST8).	n/a
pAvX	1-17,256bp	n/a	Novel plasmid; homology to other MGE limited to individual ORFs (36-91 % nucleotide identity to genes in multiple bacterial species).	**Thiol protease ScpA:** • 99.5 % amino acid identity to ScpA (GenBank AB071596) previously identified among chicken isolates from Japan (Takeuchi *et al.*, 1999, Takeuchi *et al.*, 2002). • Suggested role in poultry dermatitis (Kuramasu *et al.*, 1967). **Lysophospholipase:** • 42 % amino acid identity to a lysophospholipase encoded by *Bacillus clausii*. • Bacterial phospholipases are known virulence factors implicated in disease pathogenesis (Schmiel & Miller, 1999).
pAVY	1-1,442bp	n/a	Two genes identified with no homology to any sequence in the NCBI GenBank database.	n/a

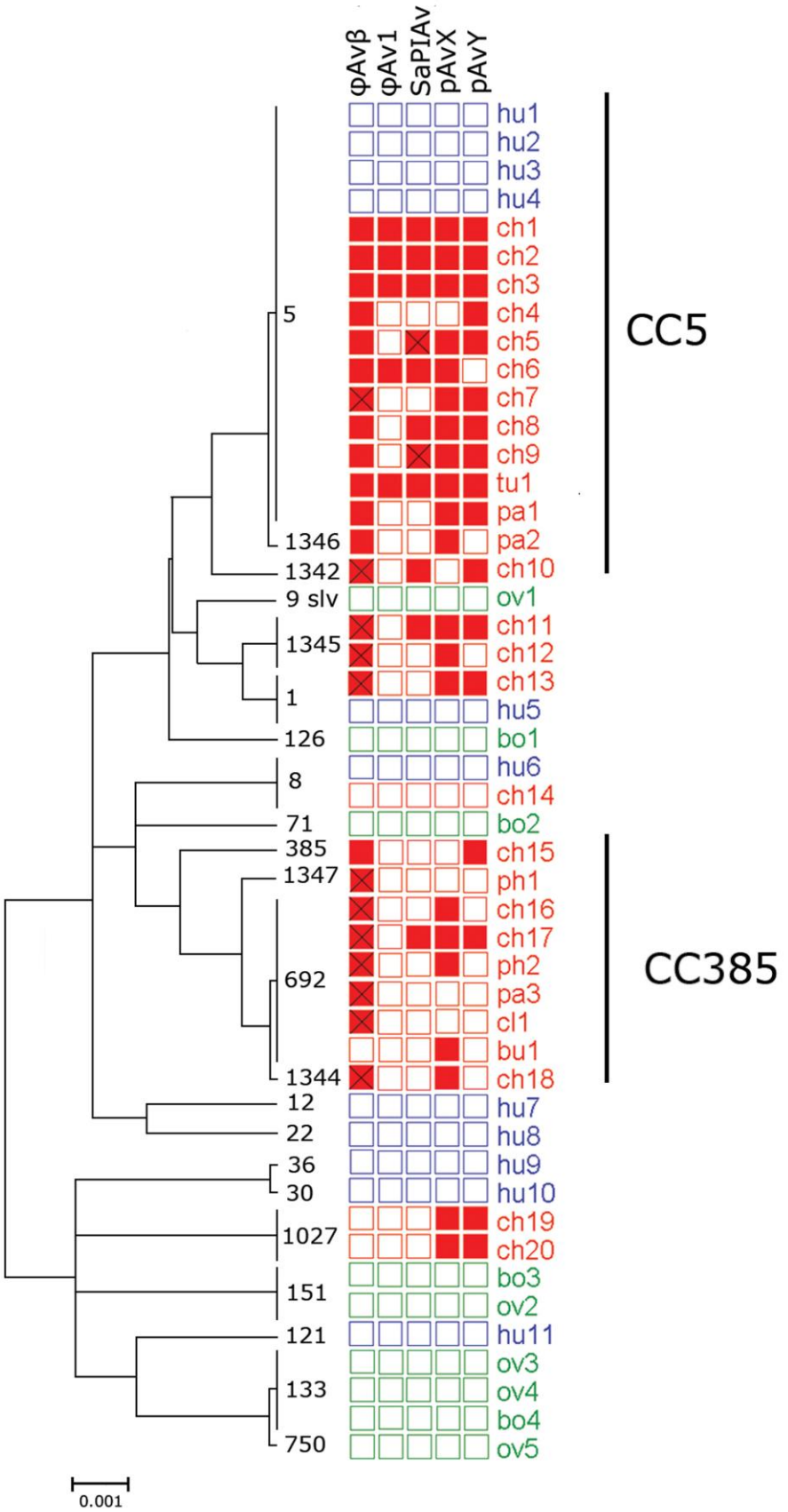

φAvβ
φAv1
SaPIAv
pAvX
pAvY
hu1
hu2
hu3
hu4
ch1
ch2
ch3
ch4
ch5
ch6
ch7
ch8
ch9
tu1
pa1
pa2
ch10
ov1
ch11
ch12
ch13
hu5
bo1
hu6
ch14
bo2
ch15
ph1
ch16
ch17
ph2
pa3
cl1
bu1
ch18
hu7
hu8
hu9
hu10
ch19
ch20
bo3
ov2
hu11
ov3
ov4
bo4
ov5
5
1346
1342
9 slv
1345
1
126
8
71
385
1347
692
1344
12
22
36
30
1027
151
121
133
750
CC5
CC385
0.001

Figure 12. The poultry CC5 clade has acquired novel MGE from an avian-specific accessory gene pool. The distribution of novel MGE identified in the ED98 genome sequence among a panel of diverse isolates of avian, human, bovine and ovine origin. Neighbour joining tree with bootstrapping consensus inferred from 500 replicates, was constructed using concatenated sequences of *S. aureus* STs from different host species including representatives of the major clonal complexes identified in the *S. aureus* species. Labels at branch tips denote STs of strains. A filled box indicates presence of an MGE and an empty box its absence, as determined by MGE-specific PCR. A crossed box indicates the presence of a related but non-identical genetic element (determined by successful amplification in 1 of 2 PCRs). Red, blue, and green boxes denote strains of avian, human, and ruminant (bovine or ovine) origin, respectively. Strain prefixes denote the host species origin of the strains including *ch*, broiler chicken; *ph*, farmed pheasant; *tu*, farmed turkey; *pa*, farmed partridge; *cl*, poultry layer; *bu* wild buzzard; *hu*, human; *bo*, cattle; and *ov*, sheep, respectively (Table 3).

Plasmid pAvX

The largest plasmid at 17 kbp, pAvX encodes a thiol protease (ScpA), previously implicated in the pathogenesis of poultry *S. aureus* infections (Takeuchi *et al.*, 1999, Takeuchi *et al.*, 2002). Also present is a gene encoding a putative lysophospholipase displaying 42 % amino acid identity to a lysophospholipase encoded by *Bacillus clausii*. pAvX was present in 11 of 13 avian CC5 strains examined and in 5 of 9 CC385 strains (Fig. 12), indicating HGT between these two major avian clades.

Plasmid pAvY

A second plasmid, pAvY, is only 1.4 kbp in size and encodes 2 hypothetical proteins of unknown function. This is also present in the two major clades, being distributed among 11 of 13 avian CC5 strains examined and 2 of 9 CC385 isolates (Fig. 12).

Staphylococcal pathogenicity island SaPIAv

ED98 carries a novel member of the SaPI family, SaPIAv. The core region is conserved with other SaPI, such as those in strains COL and RF122 (Fig. 13). However, genes for novel hypothetical proteins are present in accessory regions of the SaPI genome (Fig. 5), including 2 that encode putative novel virulence genes in the region typically occupied by superantigen genes (Table 7; (Yarwood & Schlievert, 2003). These do not display significant homology to known genes following comparison with the Genbank database. However, one (SAAV_0806) contains a signal peptide and one transmembrane helix, and the other (SAAV_0810) contains a signal peptide and four transmembrane helices, suggesting a putative role in host-pathogen interaction.

SaPIAv was present in 9 of 13 avian CC5 isolates and in 2 isolates from unrelated lineages (ST1345 and CC385). Though it is not widespread in the avian population HGT has occurred between 3 unrelated lineages (Fig. 12).

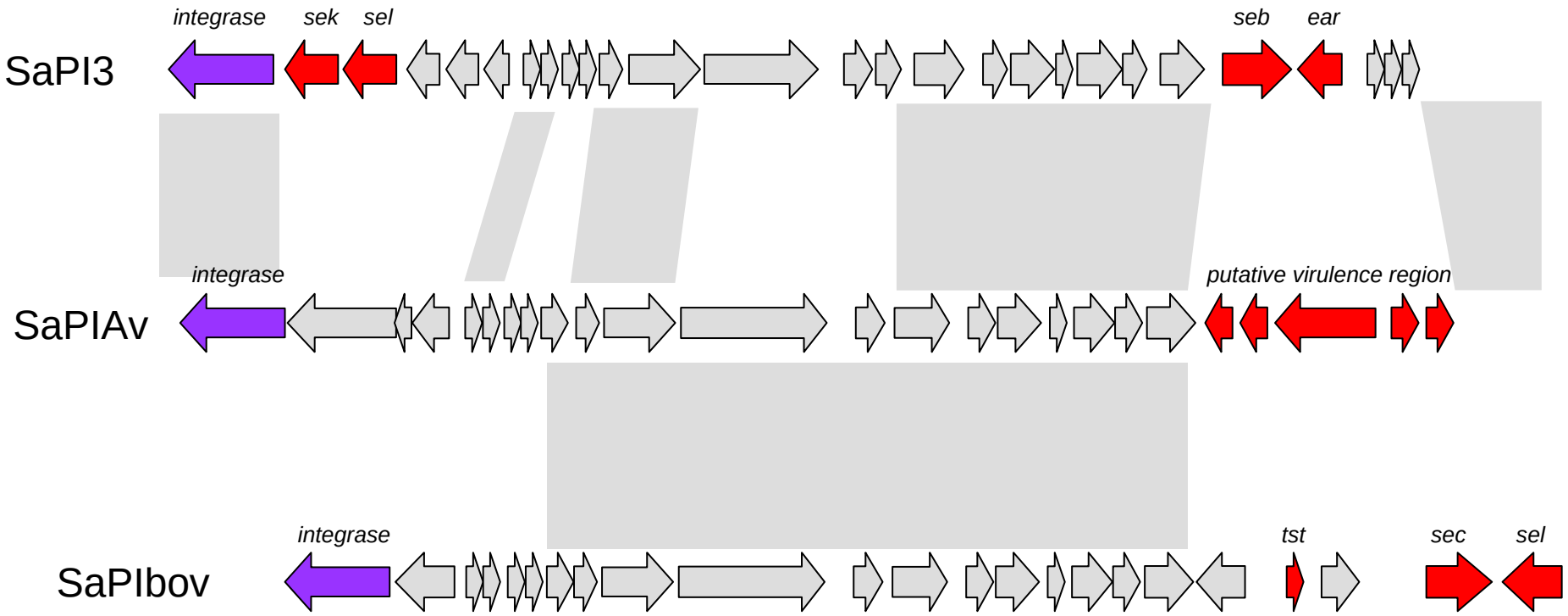


Figure 13. Novel SaPI (SaPIAv) has a conserved core and a region of novel hypothetical proteins with a putative role in virulence. The integrase sequence and genome integration point are shared with that of a SaPI3 in human isolate COL, and a large central region of the SaPI is homologous to that of SaPIbov in bovine isolate RF122. Dark grey arrows indicate SaPI ORFs, red arrows indicate virulence or putative virulence ORFs, purple arrows indicate integrase genes. Light grey boxes indicate conservation between two regions.

Prophage φAv1

A second prophage was present in ED98, designated φAv1. The phage genome has a mosaic structure typical of these MGE, in that there is one large region with homology to 77-like phage (>90 % nucleotide identity across region of 20 kbp with φPVL), and a second region encoding holin, amidase and an Ear-like protein, which is homologous (>95 % nucleotide identity) to a region in other distinct phage including φ71 and prophages in *S. aureus* strains JH1 and JH9. The *ear*-like gene (SAAV_0875) encodes a protein with putative β-lactamase activity (Yarwood *et al.*, 2002).

This prophage was found among 5 of 13 poultry CC5 isolates examined but not among other avian isolates or isolates of human or ruminant origin (Fig. 12).

4.4.3. The poultry CC5 clade has diversified from its human progenitor strain by the loss of function of genes involved in human disease pathogenesis.

In addition to the horizontal acquisition of genes by the poultry CC5 clade, examination of the genome sequence of poultry strain ED98 and comparison to its human basal strain MR1 revealed 197 SNPs, which must have arisen since divergence from a common ancestor. Though many of these are likely to be the result of genetic drift, others may have adaptive significance. Non-synonymous substitutions are observed in genes encoding a wide range of proteins, from components of transporter systems, through those involved in metabolism of a variety of substrates, to surface expressed proteins.

Of note, several of these mutations lead to a premature stop codon or frameshift so that the protein is no longer produced (Table 8). Mutations that lead to loss of function have occurred in a number of genes in strain ED98 (Table 8), including those encoding a putative surface-exposed lipoprotein, a protease SspA involved in the maturation of a second protease SspB (Massimi *et al.*, 2002), and Asp1 required for the surface expression of SraP, a cell wall-associated protein (Siboo *et al.*, 2008), in addition to the

Table 8. Distribution of pseudogenes identified in the poultry *S. aureus* strain ED98. A total of 10 pseudogenes are specific to poultry CC5 strains indicating that they arose since the most recent common ancestor with the human strain MR1; 7 were shared with human MR1 but intact in N315 indicating they arose in the lineage ancestral to the poultry clade and human MR1 but since divergence with N315; and 22 pseudogenes in ED98 are shared with the human strains MR1 and N315 (GenBank BA000018), indicating loss of function prior to the most recent common ancestor of the entire CC5 radiation (Fig. 10).

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
spA	Staphylococcal Protein A	Inhibition of opsonophagocytosis by non-specific binding to human IgG (37, 38) adherence to human lung epithelial cells via the TNFR1 receptor (39), and adhesion to human platelets by binding of von Willebrand factor (40, 41)	Pseudogene	Intact	Intact	19/19
asp1	Accessory secretory protein 1	Required for expression of SraP, a cell-wall-associated protein involved in adhesion to human platelets and implicated in pathogenesis of infective endocarditis (42)	Pseudogene	Intact	Intact	10/19
sspA	Staphylococcal serine protease	Required for maturation of SspB, a serine protease that interacts with human host factors such as fibrinogen to alter structure of blood clots (43)	Pseudogene	Intact	Intact	5/19
SAAV_1787	RNA polymerase sigma-70	Regulation of gene transcription	Pseudogene	Intact	Intact	19/19

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
SAAV_0389	Conserved hypothetical protein in VSaα	Unknown	Pseudogene	Intact	Intact	16/19
SAAV_1594	Conserved hypothetical protein, Fic protein family	Putative cell division role	Pseudogene	Intact	Intact	19/19
SAAV_0046	Putative lipoprotein	Host-pathogen interactions	Pseudogene	Intact	Intact	19/19
SAAV_0047	Integrase (transposon copy 1)	Mobility of transposon	Pseudogene	Not present	Not present	n/d
SAAV_1689	Integrase (transposon copy 2)	Mobility of transposon	Pseudogene	Not present	Not present	n/d
SAAV_1941	Integrase (transposon copy 3)	Mobility of transposon	Pseudogene	Not present	Not present	n/d
SAAV_0217	Hexitol dehydrogenase	Metabolism	Pseudogene	Pseudogene	Intact	n/d
SAAV_0556	Putative pyridine nucleotide-disulfide oxidoreductase	Metabolism	Pseudogene	Pseudogene	Intact	n/d

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
SAAV_0937	2-isopropylmalate synthase	Metabolism	Pseudogene	Pseudogene	Intact	n/d
SAAV_1830	Conserved hypothetical protein in VSaβ	Unknown	Pseudogene	Pseudogene	Intact	n/d
SAAV_2323	Transporter (AcrB/AcrD/AcrF family)	Putative multidrug efflux transporter	Pseudogene	Pseudogene	Intact	n/d
SAAV_2728	Conserved hypothetical protein	Unknown	Pseudogene	Pseudogene	Intact	n/d
SAAV_0145	NAD-dependant formate dehydrogenase	Metabolism	Pseudogene	Pseudogene	Intact	n/d
SAAV_1795	Transposase	Mobility of MGE	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_1813	Transposase	Mobility of MGE	Pseudogene	Pseudogene	Pseudogene	n/d
hlb	β-hemolysin	Sphingomyelinase activity (44)	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_0031	Conserved hypothetical protein (thiolase active site)	Putative metabolic role	Pseudogene	Pseudogene	Pseudogene	n/d

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
Gene not annotated	Conserved hypothetical protein (1 signal peptide)	Putative membrane transporter	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_0999	Glycosyl transferase, group1	Carbohydrate modification of the cell envelope	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_1592	Putative enterotoxin	Cytotoxicity	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (copy 1)	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (copy 2)	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (copy 3)	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (copy 4)	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
cysS	Cysteinyl-tRNA synthetase	Catalyses attachment of amino acid to transfer RNA molecule	Pseudogene	Pseudogene	Pseudogene	n/d
yent 1 & 2	Superantigen	Secreted toxin with mitogenic activity	Pseudogene	Pseudogene	Pseudogene	n/d

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
Gene not annotated	Conserved hypothetical protein (copy 5)	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_1824	Epidermin immunity protein in VSaB	Lantibiotic resistance	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein in VSaβ (signal peptide & iron-sulphur binding site)	Putative membrane transporter and/or metabolic role	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (signal peptide & iron-sulphur binding site)	Putative membrane transporter and/or metabolic role	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein (signal peptide & iron-sulphur binding site)	Putative membrane transporter and/or metabolic role	Pseudogene	Pseudogene	Pseudogene	n/d
Gene not annotated	Conserved hypothetical protein	Unknown	Pseudogene	Pseudogene	Pseudogene	n/d
SAAV_0240	NADH dehydrogenase subunit	Metabolism	Pseudogene	Pseudogene	Pseudogene	n/d

Gene/locus	Gene product	Function/putative function	Integrity of gene in specified strain			Presence in 19 CC5 poultry isolates
			ED98	MR1	N315	
SAAV_2204	Hypothetical protein, similar to transposase for IS232	Mobility of MGE	Pseudogene	Pseudogene	Pseudogene	n/d
oppB	Peptide ABC transporter, permease protein	Transmembrane transport protein	Pseudogene	Pseudogene	Pseudogene	n/d

previously identified *spA* pseudogene.

Ten pseudogenes are present in ED98 but not in basal human isolate MR1, indicating that they arose since the most recent common ancestor of these two strains. Seven are shared with MR1 but are intact in N315, indicating that they arose in the lineage ancestral to the poultry clade and human MR1, but since divergence with N315. Finally, 22 pseudogenes in ED98 are shared with the human strains MR1 and N315, indicating loss of function prior to the most recent common ancestor of the entire CC5 radiation (Table 8).

PCR and sequencing of 19 CC5 poultry isolates was performed to determine whether the 10 pseudogenes unique to ED98 are found throughout the CC5 poultry clade. Several were found in all 19, including *spA*, a putative regulatory gene SAAV_1787 and a putative lipoprotein (Table 8). Each pseudogene was present in at least five of the 19 isolates, confirming that all pseudogenes are present in multiple branches of the poultry ST5 clade and are not unique to ED98.

4.5. Discussion

4.5.1. Genome evolution in the CC5 lineage

In order to investigate the genomic basis for adaptation of the CC5 poultry clade to its adopted host, the whole genome sequence of poultry isolate ED98 was compared to the genomes of other *S. aureus* strains, including basal human strain MR1. Results demonstrate that evolution of poultry strain ED98 involved acquisition of novel genomic regions not found in the CC5 human associated progenitor MR1, or in other known *S. aureus* sequences. Several novel MGE were identified, including two prophage (φAv1 and φAvβ), a novel member of the SaPI family of pathogenicity islands (SaPIAv) and two plasmids (pAvX and pAvY).

Furthermore, PCR screening reveals that the majority of these MGE are present in avian isolates from multiple distinct lineages, but are conspicuously absent from human, bovine or ovine isolates (Fig. 12). This suggests that the poultry CC5 clade has acquired novel MGE from an avian-specific accessory gene pool, maintained within unrelated strains that occupy the same avian host niche.

Importantly, avian-specific MGE sequences are found in lineages represented by very few strains in the collection of poultry isolates (Chapter 3; Fig. 7). This observation supports the hypothesis that these strains represent other groups that have recently switched from a human to poultry host. Data suggest that in common with the CC5 poultry clade, these clades are undergoing host adaptation, at least in part through acquisition of avian-specific MGEs.

These findings also suggest that this CC5 sub-lineage, which includes the poultry strains and human isolates from Poland, may be particularly proficient at acquiring MGE. This property may have led to the emergence of both the successful poultry pathogen and the human MRSA clone predominant in Polish hospitals in the 1990s.

In addition to HGT, analyses suggest that small scale changes in the core genome may also contribute to host adaptation as SNPs were identified between strains ED98 and MR1. Protein modelling, functional characterisation and/or mutation studies are required to ascertain the effect of specific nucleotide differences leading to amino acid changes. Importantly, 10 of these mutations lead to the formation of pseudogenes in ED98. This is either due to substitution that causes a premature stop codon, or to insertion of one or a number of bases that introduce frameshift. Seven pseudogenes arose in the lineage ancestral to the poultry clade and human MR1, but since divergence with N315, whilst a further 22 arose prior to the most recent common ancestor of the entire CC5 radiation, suggesting that genome reduction has had a part to play both in evolution of the CC5 lineage of *S. aureus* as a whole, and in particular in adaptation to a poultry niche. Notably, several of the pseudogenes that have arisen in the poultry clade

encode proteins involved in human disease pathogenesis. These include a putative surface-exposed lipoprotein, proteins involved in production of virulence factors SspB and SraP (Massimi *et al.*, 2002, Siboo *et al.*, 2008) and, as was previously observed in phenotypic studies of human and poultry associated strains, SpA (Foster, 2005, Gemmell *et al.*, 1991, Gomez *et al.*, 2006, Hartleib *et al.*, 2000, O'Seaghdha *et al.*, 2006).

Identical nonsense mutations that inactivate *spA*, the regulatory gene SAAV_1787 and a putative lipoprotein were found in all poultry CC5 strains examined, indicating that they happened early in the evolution of the poultry clade. All of the pseudogenes were present in at least five of the 19 isolates, suggesting that they each play a role in adaptation to a novel niche as they have arisen in one strain and been maintained in multiple branches of the CC5 poultry lineage.

Genome reduction is characteristic of bacterial species undergoing niche restriction (Eppinger *et al.*, 2006, Thomson *et al.*, 2008), and in this case may be due to different requirements of the avian host environment. Proteins previously important for pathogenesis may no longer be of use as they do not interact in the same way (or at all) with their novel host. Selective pressure leads to loss of the gene because of the fitness cost of producing such a protein, or because there are deleterious effects of interaction with the new host.

4.5.2. Avian host adaptation

The genome of CC5 poultry strain ED98 has been shaped by the loss of function of genes involved in human disease pathogenesis and by HGT. The identification of several novel MGE which are widely distributed among avian isolates from distinct clonal lineages, and which are not associated with human or ruminant strains, indicates the presence of an avian niche-specific accessory gene pool for *S. aureus*. The maintenance of such a gene pool implies that it confers an important selective advantage to bacteria residing in the niche. Genes encoded on these MGE may be important for

colonisation and/or invasion of an avian host, whose biology differs from that of a mammalian host in several recognised ways.

The chicken genome sequence, published in 2004, revealed major differences in structure and gene content between chicken and mammalian genomes, highlighting the features to which the CC5 poultry clade had to adapt following a host switch. Notably, only around 60 % of chicken protein-coding genes have a single human orthologue (International Chicken Genome Sequencing Consortium, 2004). Of these, genes involved in host defence exhibit greater divergence from the human form than, for example, those involved in cytoplasmic and nuclear functions (International Chicken Genome Sequencing Consortium, 2004). Furthermore, manual scrutiny of the genome revealed that some of the remaining 40 % of genes encode probable immune proteins that have diverged to such an extent that orthologue searches do not detect them (International Chicken Genome Sequencing Consortium, 2004).

Other studies have identified specific examples of differences in the immune systems of mammals and birds. Toll-like receptors (TLRs) are a key component of the innate immune system as they are involved in microbe recognition (Takeuchi & Akira, 2010), but homology between chicken TLRs and those of mice or humans can be as low as 28 % in some families (Yilmaz *et al.*, 2005). Importantly, analysis has shown that sequence changes in the ligand-binding domains of TLR1 in chickens are a principal component of host evolution (Yilmaz *et al.*, 2005), suggesting that there may be strong selective pressure for corresponding changes in microbes, in order that they evade detection. In another example, avian cytokines perform the same function in both mammals and birds in that they act as immuno-regulatory molecules, but studies have shown that avian cytokines have a remarkably different structure to their mammalian counterparts (for review see Staeheli *et al.*, 2001). Structural variation is also observed in immunoglobulins. The avian equivalent of human IgG is IgY, which has a structurally distinct Fc region which, unlike that of IgG, does not bind the major *S. aureus* surface protein SpA (Barkas & Watson, 1979, Warr *et al.*, 1995).

It seems likely that such differences in host defences present a major challenge to the *S. aureus* CC5 poultry clade. However, there are other important aspects of avian physiology to be considered. Comparative genomic analysis demonstrated that orthologues of mammalian genes encoding salivary-associated proteins are entirely absent from the chicken genome (International Chicken Genome Sequencing Consortium, 2004), whilst proteins with scavenger receptor cysteine rich (SRCR) domains are significantly over-represented, a feature thought to be linked to mucosal homeostasis (Kang & Reid, 2003). This may be important to *S. aureus* as it frequently colonises mucosal surfaces such as the anterior nares, and as such may be adapted to specific conditions unique to humans.

Another notable difference is variation in body temperature; normal avian body temperature is 42°C whereas human is 37°C. Features of host metabolism may also affect the environment inhabited by bacteria, such as differences in excretion of nitrogenous waste. Whereas mammals excrete urea, birds excrete uric acid, probably to avoid the build up of soluble urea during development in the egg (International Chicken Genome Sequencing Consortium, 2004).

It is highly likely that some of the mutations observed in ED98 and/or some of the genes of unknown function located on novel MGE are involved in adaptation to such changes in host environment. For example, the prophage φAvβ carries an ornithine cyclodeaminase gene, which may relate to adaptation to an avian environment with a high concentration of ornithine due to differences in processing of nitrogenous waste. Of note, the amino acid proline is important to many bacteria as an osmoprotectant and antioxidant, and arginine is utilised in its generation (Townsend *et al.*, 1996). Usually, arginine is transported into the *S. aureus* cell via an arginine/ornithine antiporter (ArcD) but if there is a high extracellular concentration of ornithine, bacteria may instead efflux arginine and take up ornithine. If this occurs, the usual means for *S. aureus* to generate proline (*i.e.* via arginine) may be impaired. Thus, the single-step conversion of ornithine

to proline, a process that involves ornithine cyclodeaminase, could provide an advantage in such an environment.

Loss of the *sspA* gene, encoding Staphylococcal serine protease SspA, may also be related to host adaptation. This protein is required for maturation of SspB, a second serine protease that interacts with fibrinogen α and β chains (altering the structure of blood clots), and also with fibronectin and kininogen. The chicken homologues of fibrinogen α, fibrinogen β, fibronectin and kininogen exhibit only 44 %, 65 %, 81 % and 35 % homology, respectively, with the human form. Proteolytic activity of SspB requires specific target sequences that may not be conserved in an avian host.

There are also several genes encoded on MGE that may be involved in persistence, virulence or resistance, contributing to the emergence of this CC5 sublineage as a major global poultry pathogen. The pathogenicity island SaPIAv carries two hypothetical proteins with transmembrane helices and signal peptides, which may suggest a role as membrane transporters. Plasmid pAvX carries a putative lysophospholipase and prophage φAvβ has a putative membrane protease, both of which are from protein families implicated in pathogenesis in several bacterial species (Dubin, 2002, Schmiel & Miller, 1999). Finally, φAv1 has an Ear-like protein found on several other phage, with 80 % homology to the protein Ear, a putative β-lactamase protein found on SaPI1, SaPI3, SaPI5 and SaPImw2.

4.5.2. Conclusion

In conclusion, these data have revealed genetic evidence for the host adaptation of the poultry clade of the CC5 lineage through extensive acquisition of MGEs from other *S. aureus* isolates colonising poultry, and through loss of gene function. Previous population genetic analysis of the poultry clade indicated a human-to-poultry host jump and intercontinental dissemination of the poultry adapted clone, and data presented here reveal the underlying genomic changes associated with this.

5. Genome-wide analysis of positive selection in *S. aureus*

5.1. Introduction

In order to survive, bacterial pathogens must adapt to selective pressures such as the immune syatem, inhibition by therapeutics and changes in host biology. Positive selection leading to fixation of beneficial mutations is an important facet of adaptive evolution. In recent years, phylogenetic methods have been applied to many taxa to identify genes under positive selection by comparing rates of synonymous versus non-synonymous mutation (dN/dS or ω) (for example, humans, chimpanzees and mice (Clark *et al.*, 2003, Nielsen *et al.*, 2005), *Drosophila* (Richards *et al.*, 2005) and *E. coli* (Petersen *et al.*, 2007)). In particular, studies on a genome-wide scale are a powerful means to identify proteins that are adapting to selective forces in all strains, and in certain niches only.

Population genomic analysis of avian isolates of *S. aureus* in the current study has revealed the importance of gene acquisition and loss in host adaptive evolution. However, such processes are not the only mechanisms of genome evolution. This component of the project aims to assess the relative importance of positive selection in the niche adaptation of *S. aureus*.

5.2. Aim

To perform genome-wide analysis of positive selection in 30 strains of *S. aureus* representing diverse evolutionary lineages and host associations, in order to identify genes undergoing adaptive evolution.

5.3. Materials and methods

Python script preparation and recombination analysis were performed by Jon Bollback (The University of Edinburgh).

5.3.1. Genome sequence retrieval and identification of orthologous genes

Genome sequences for 30 strains of *S. aureus* were downloaded from the NCBI database, the EMBL database or the Sanger Institute website, or were generated by authors of this study and collaborators (Table 9). Of these, 16 genomes were available as coding sequences (CDS) (.ffn files) as annotation had been performed previously, whilst the remaining 14 were available in the form of multiple contiguous sequences (contigs) assembled from shotgun sequencing reads. Contigs were concatenated and submitted to the RAST annotation server (Aziz *et al.*, 2008) for prediction of CDS. CDS spanning overlaps between contigs were manually trimmed where a gene fragment was detected on one or both contig edges, using Artemis software (Rutherford *et al.*, 2000). CDS for all 30 genomes were concatenated into a single document, which was formatted to create a BLASTn database.

A python script was written to identify groups of orthologous genes. Briefly, all CDSs from *S. aureus* strain COL were aligned against the database of concatenated genomes using BLAST (e-value cut off 0.00001). If a match for a given COL gene was identified in one of the other 29 genomes, and a reciprocal BLAST of the matching CDS returned the original COL CDS, the genes were considered orthologous and the protein sequence was retrieved. Orthologous protein sequences were then aligned with ClustalW (Larkin *et al.*, 2007) and protein sequences replaced by the corresponding nucleotide sequences, preserving the gaps obtained during protein alignment. Orthologue groups containing sequences from at least 3 of the 30 genomes were included for further analyses.

5.3.2. Tree construction and test for positive selection

Phylogenetic trees were inferred for each orthologue group using the PAUP* software package (Wilgenbusch & Swofford, 2003). In order to test for positive selection, the

Table 9. Strains of *S. aureus* considered in a genome-wide analysis to identify genes displaying evidence for positive selection.
Genome sequences were either publically available or sequenced by authors and/or collaborators

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
ED98	Poultry skeletal infection isolate	ST5 (CC5)	N/A	Complete	NCBI (CP001781)	Univ. of Edinburgh	This study
ED133	Ovine clinical mastitis isolate	ST133 (CC133)	N/A	Complete	NCBI (CP001996)	Univ. of Edinburgh	Guinane *et al.*, 2010
DL190	Rabbit isolate	ST121 (CC121)	N/A	Complete	Unpublished	Inst. Valenciano de Investigaciones agrarias	Unpublished
RF122	Bovine milk isolate	ST151 (CC705)	N/A	Complete	NCBI (AJ938182.1)	Univ. of Minnesota	Herron-Olson *et al.*, 2007

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
Mu3	Human VISA clinical isolate	ST5 (CC5)	HA	Complete	NCBI (AP009324.1)	Juntendo Univ., NITE	Neoh *et al.*, 2008
JH1	Human MRSA isolate	ST5 slv (CC5)	HA	Complete	NCBI (CP000736.1)	Rockefeller Univ., JGI	Mwangi *et al.*, 2007
Newman	Early human clinical isolate	ST8 (CC8)	CA	Complete	NCBI (AP009351.1)	Juntendo Univ., NITE	Baba *et al.*, 2008
JH9	Human MRSA & VRSA isolate	ST5 slv (CC5)	HA	Complete	NCBI (CP000703.1)	Rockefeller Univ., JGI	Mwangi *et al.*, 2007
NCTC8325	Early human clinical isolate	ST8 (CC8)	Unknown	Complete	NCBI (CP000253.1)	Univ. of Oklahoma	Gillaspy *et al.*,2006

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
COL	Early human hospital-acquired MRSA (1960s)	ST250 (CC8)	HA	Complete	NCBI (CP000046.1)	J Craig Venter Institute	Gill *et al.*, 2005
MSSA476	Human MSSA clinical isolate	ST1 (CC15)	CA	Complete	NCBI (BX571857.1)	Sanger Institute	Holden *et al.* 2004
MRSA252	EMRSA-16 clone prevalent in UK human hospitals	ST36 (CC30)	HA	Complete	NCBI (BX571856.1)	Sanger Institute	Holden *et al.*, 2004
MW2	Human MRSA: septicaemia and septic arthritis	ST1 (CC15)	CA	Complete	NCBI (BA000033.2)	Juntendo Univ., NITE	Baba *et al.*, 2002
Mu50	Human VRSA clinical isolate	ST5 (CC5)	HA	Complete	NCBI (BA000017.4)	Juntendo Univ., NITE	Kuroda *et al.*, 2001

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
N315	Human MRSA clinical isolate	ST5 (CC5)	HA	Complete	NCBI (BA000018.3)	Juntendo Univ., NITE	Kuroda *et al.*, 2001
USA300 FPR3757	Major USA human CA clone	ST8 (CC8)	CA	Complete	NCBI (CP000255.1)	BCM-HGSC	An Diep *et al.*, 2006
USA300 TCH1516 (USA300-MR-HOU)	Major USA human CA clone	ST8 (CC8)	CA	Complete	NCBI (CP000730.1)	UCSF	Highlander *et al.*, 2007
TW20 (0582)	Human highly transmissible, invasive MRSA	ST239 (CC8)	HA	Complete	Sanger Institute [3]	Sanger Institute	Holden *et al.*, 2010

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
MR1	Hospital-acquired polish human MRSA isolate	ST5 (CC5)	HA	Incomplete	NCBI [1]	Univ. of Edinburgh	This study
CF-Marseille	Human cystic fibrosis associated isolate	ST5 (CC5)	Unknown	Incomplete	EMBL-EBI [2]	Univ. de la Méditerranée	Rolain *et al.*, 2009
H1374 (H174)	UK human hospital isolate	ST38 (CC30)	HA	Incomplete	Sanger Institute [4]	Sanger Institute	Enright *et al.*, 2000
NOH4 (Nottm4)	Human Epidemic MRSA strain	ST22 slv (CC22)	HA	Incomplete	Sanger Institute [5]	Sanger Institute	None
EMRSA15	1 of 2 MRSA clones prevalent in UK human hospitals	ST22 (CC22)	HA	Incomplete	Sanger Institute [6]	Imperial College London / Sanger Institute	None

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
TCH130	Human skin isolate	ST72 (CC8)	CA	Incomplete	NCBI (GPID 31541)	BCM-HGSC	None
TCH60	Human skin isolate	ST30 (CC30)	CA	Incomplete	NCBI (GPID 31539)	BCM-HGSC	None
TCH70	Human skin isolate	ST1 (CC15)	CA	Incomplete	NCBI (GPID 31543)	BCM-HGSC	None
TCH959	Human skin isolate	ST1159 (CC7)	CA	Incomplete	NCBI (GPID 19921)	BCM-HGSC	None
MN8	Human urogenital tract isolate	ST30 (CC30)	Unknown	Incomplete	NCBI [7]	BCM-HGSC	None

Strain	Isolation features	MLST (CC)	Hospital or community associated	Sequencing status	Source	Institution	Reference
JKD6008	Human blood culture isolate isolate †	ST239 (CC8)	HA	Incomplete	NCBI [8]	Monash Univ.	Howden *et al.*, 2008
JKD6009	Postsurgical wound infection isolate ‡	ST239 (CC8)	HA	Incomplete	NCBI [9]	Monash Univ.	Howden *et al.*, 2008

[1] WGS ACZQ00000000

[2] WGS CABA00000000

[3] ftp://ftp.sanger.ac.uk/pub/pathogens/sa/ Sa_0582.db

[4] ftp://ftp.sanger.ac.uk/pub/pathogens/sa/454/ H1374_denovo_454AllContigs.fna

[5] ftp://ftp.sanger.ac.uk/pub/pathogens/sa/454/ NOH4_denovo_454AllContigs.fna

[6] ftp://ftp.sanger.ac.uk/pub/pathogens/sa/ EMRSA15.dna

N/A, not applicable

[7] WGS ACJA00000000

[8] WGS ABRZ00000000

[9] WGS ABSA00000000

† Patient developed MRSA bacteraemia and endocarditis following surgery; pre- glycopeptide antibiotic treatment.

‡ Patient developed MRSA bacteraemia and endocarditis following surgery; after vancomycin therapy.

method of Nielsen and Yang (1998) was utilised, as implemented in PAML (Yang, 1997) and described previously (Petersen *et al.*, 2007).

The method tests for the presence of positively selected sites by allowing ω to vary among sites. Data were fitted to alternative models of selection, either positive or none, by comparing M1a with M2a, and M7 with M8 (Yang, 1997, Yang *et al.*, 2000). Genes were considered to show evidence for positive selection across all strains if they matched both models M2a and M8, following fdr correction. Aligned nucleotide sequences for each orthologue group exhibiting evidence for selection were manually inspected to identify alignment errors.

Positive selection acting on one lineage only is indicated by an elevated ω in this lineage. Analysis was performed using codeml, as implemented in the PAML package (Yang, 1997). Rates were compared between human-associated and non-human animal-associated lineages with a Bonferroni multiple test correction.

5.3.3. Test for recombination

Recombination can interfere with the molecular signal for positive selection as the method used for detecting positive selection assumes that all sites share the same underlying phylogeny (Anisimova & Liberles, 2007, Shriner *et al.*, 2003). Therefore, genes that showed evidence for positive selection were analysed for evidence of recombination, using a statistical test implemented in HyPhy (Pond *et al.*, 2005) as described previously (Petersen *et al.*, 2007).

5.3.4. Multi Locus Sequence Typing and phylogenetic analysis

Sequence Types (ST) of strains included in the analysis were either previously published or were determined from the published genome sequence. For each of the seven housekeeping genes employed for *S. aureus* MLST, the nucleotide sequence of a single allele was downloaded from the MLST website (http://saureus.mlst.net/). Relevant regions of each genome were identified using BLASTn sequence comparison visualized

using the Artemis Comparison Tool (Carver *et al.*, 2005) and sequences were submitted to the MLST website to determine allele numbers and resulting ST. Reconstruction of evolutionary relationships was carried out using concatentated MLST sequence data, as described in materials and methods.

5.3.5. Gene annotation

Protein function was inferred from existing annotations of strain COL including gene name, homology to PFAMs and COG assignment, and was supplemented by additional BLAST searches. Assignment of encoded proteins to functional group was performed for (i) all genes identified in orthologue search, and (ii) genes displaying evidence for positive selection. In order to detect statistically significant enrichment of genes under selection in a particular functional group, numbers for each group were compared using a χ^2 test (1 degree of freedom).

5.4. Results

5.4.1. Comparative analysis of the *S. aureus* genome

In order to identify genes with orthologues in multiple lineages of *S. aureus*, comparative genomic analysis was performed with 30 genome sequences for strains of diverse origin (Table 9). This included bovine (RF122), ovine (ED133), avian (ED98) and rabbit (DL190) strains, in addition to 26 isolates of human origin. Of the human isolates, which included MRSA, MSSA, VISA and VRSA strains, 14 were hospital-acquired and 9 were community-acquired (status was unknown for 3 strains) including isolates associated with invasive disease and commensalism. Furthermore, the strains analysed represent a cross section of the genetic diversity across the *S. aureus* species, as determined by MLST (Fig. 14).

Sets of orthologous gene sequences were identified by reciprocal BLAST analysis. This revealed a *S. aureus* genome of 2602 genes, comprising 2055 (79 %) core genes

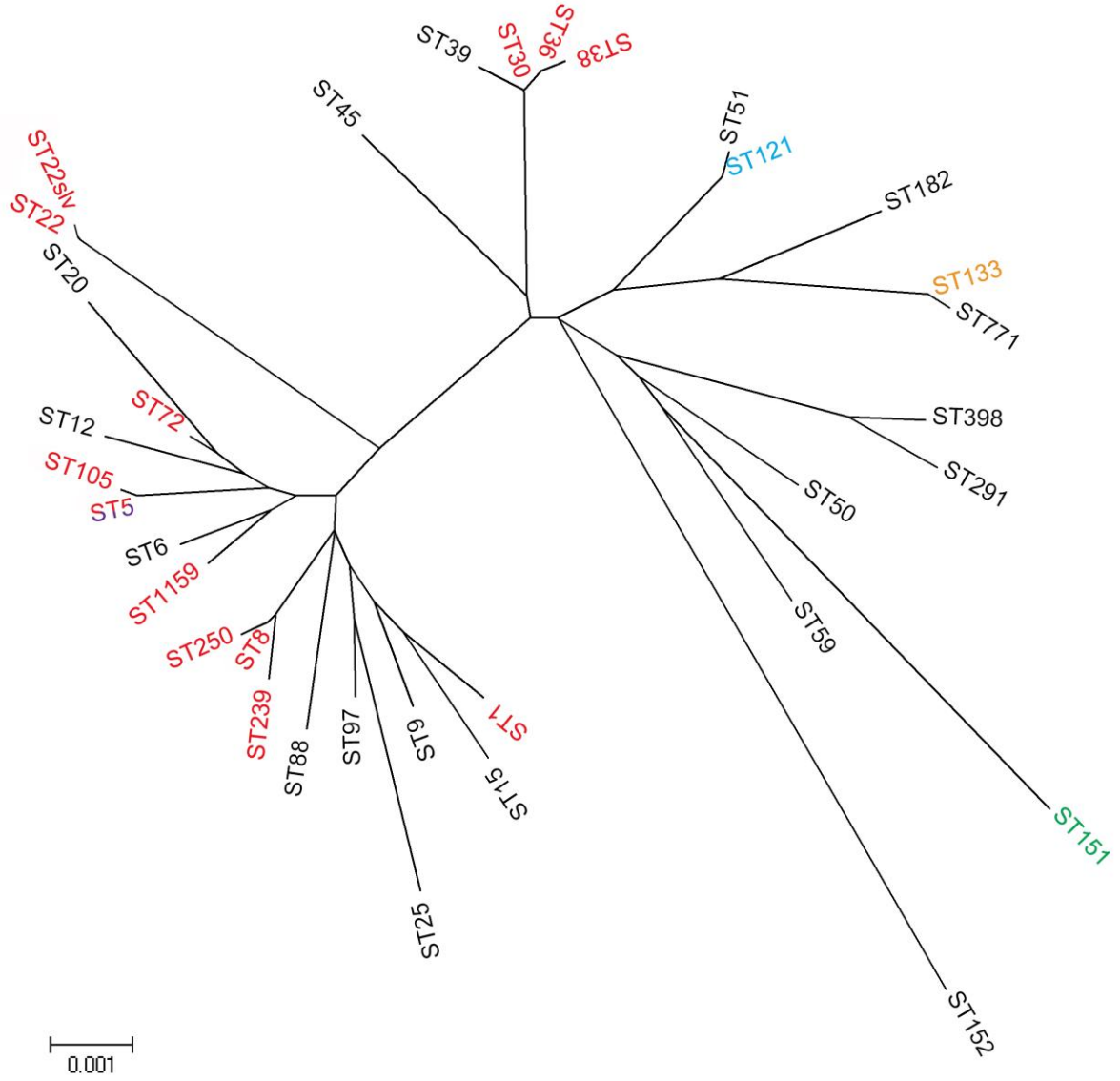


Figure 14. *S. aureus* isolates analysed in this study represent breadth of diversity in the species. Strains analysed are from distinct environmental niches, and phylogenetic analysis of 7 concatenated MLST alleles demonstrates that they represent divergent evolutionary lineages. Coloured STs indicate lineages represented by strains in the study: red are human isolates, green is bovine, orange is ovine, blue is rabbit, red/purple is a lineage with both human and chicken isolates. Black STs represent remaining major lineages (CCs) identified in *S. aureus* species.

(orthologues in at least 28 of the 30 genomes) and 546 (21 %) accessory genes (orthologues in 27 or less of the 30 genomes). It should be noted that 12 of the genomes contained gaps between contigs. Such gaps are frequently located in genes with internal repeats or regions with high homology to other regions of the genome, including genes encoding the MSCRAMM family of surface proteins. Thus, in a small number of cases, genes designated accessory rather than core genes may reflect this limitation, rather than narrow distribution of the gene across the *S. aureus* species. Furthermore, the approach used to identify orthologues depends upon a gene being present in the genome of COL, the reference strain. Therefore, the study excludes many genes of the accessory or core-variable genome, and should thus be considered predominantly an analysis of positive selection occurring in the *S. aureus* core genome.

5.4.2. Genome-wide positive selection drives *S. aureus* genome evolution

In order to assess the contribution of positive selection to evolution of the *S. aureus* genome, gene sequences with orthologues in multiple genomes were analysed for the molecular signature of positive selection based upon dN/dS (ω). In total, 125 genes representing 4.8 % of the complete *S. aureus* genome (which is defined as the 2602 genes identified in the orthology search) showed significant evidence for positive selection in the absence of recombination (Tables 10-12), and are distributed throughout the genome (Fig. 15).

Genes under positive selection were found in all functional categories except for intracellular trafficking and secretion, cell motility and secretion, and chromatin structure and dynamics, which make up a small proportion of the *S. aureus* genome (1.4 %, Fig. 16). Of note, 27 % (n=34) of the genes under selection encode hypothetical proteins of unknown function.

Table 10. Genes displaying evidence for positive selection, with elevated rate of non-synonymous mutations across all branches of the *S. aureus* species phylogenetic tree examined

ID	Gene name	Product	COG	Functional group
SACOL0422	-	ABC transporter ATP-binding protein	COG1131V	Defence mechanisms
SACOL2646	-	DNA-binding response regulator	COG0745TK	Signal transduction mechanisms, Transcription
SACOL1038	-	Membrane protein	COG4652S	Function unknown
SACOL1500	-	Hypothetical protein	COG2738R	General function prediction only
SACOL0690	-	ABC transporter ATP-binding protein	COG1121P	Inorganic ion transport and metabolism
SACOL0705	-	Iron compound ABC transporter permease protein	COG0609P	Inorganic ion transport and metabolism
SACOL2165	-	Iron compound ABC transporter permease protein	COG0609P	Inorganic ion transport and metabolism
SACOL0438	*ssb2*	Single-stranded DNA-binding protein	COG0629L	DNA replication, recombination and repair
SACOL0627	*ung*	Uracil-DNA glycosylase	COG0692L	DNA replication, recombination and repair
SACOL1382	*sbcC*	Exonuclease SbcC	COG0419L	DNA replication, recombination and repair
SACOL1390	*parC*	DNA topoisomerase IV subunit A	COG0188L	DNA replication, recombination and repair
SACOL1671	-	Holliday junction resolvase-like protein	COG0816L	DNA replication, recombination and repair
SACOL1955	*dinP*	DNA-damage-inducible protein P	COG0389L	DNA replication, recombination and repair

ID	Gene name	Product	COG	Functional group
SACOL2147	-	Transcriptional antiterminator Bglg family/DNA-binding protein	COG3711K	Transcription
SACOL1688	-	D-tyrosyl-tRNA deacylase	COG1490J	Translation, ribosomal structure and biogenesis
SACOL2739	*rnpA*	Ribonuclease P	COG0594J	Translation, ribosomal structure and biogenesis
SACOL0463	-	Hypothetical protein	COG2918H	Coenzyme metabolism
SACOL0460	*guaB*	Inosine-5'-monophosphate dehydrogenase	COG0516F, COG0517R	Nucleotide transport and metabolism
SACOL1655	*mtn*	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	COG0775F	Nucleotide transport and metabolism
SACOL1476	-	Amino acid permease	COG0531E	Amino acid transport and metabolism
SACOL1563	*lpdA*	2-oxoisovalerate dehydrogenase E3 component lipoamide dehydrogenase	COG1249C	Energy production and conversion
SACOL0119	-	Putative cell wall surface anchor family protein	Unknown	Unknown
SACOL0358	-	Hypothetical protein	Unknown	Unknown
SACOL0393	-	Hypothetical protein	Unknown	Unknown

ID	Gene name	Product	COG	Functional group
SACOL0989	-	Hypothetical protein	Unknown	Unknown
SACOL1142	isdD	Iron compound ABC transporter protein	Unknown	Unknown
SACOL1144	*IsdF*	Iron compound ABC transporter permease protein	Unknown	Unknown
SACOL1345	-	Hypothetical protein	Unknown	Unknown
SACOL1348	-	Hypothetical protein	Unknown	Unknown
SACOL1766	-	Hypothetical protein	Unknown	Unknown
SACOL2123	-	Hypothetical protein	Unknown	Unknown
SACOL2429	-	Hypothetical protein	Unknown	Unknown
SACOL2470	-	Hypothetical protein	Unknown	Unknown
SACOL0034	*mecR1*	Methicillin-resistance MecR1 regulatory protein	COG4219KT	Transcription, Signal transduction mechanisms
SACOL0064	-	Metallo-beta-lactamase family protein	COG0491R	General function prediction only
SACOL0099	*sirA*	Iron compound ABC transporter iron compound-binding protein Sira	COG0614P	Inorganic ion transport and metabolism
SACOL0477	-	Type I restriction-modification system S subunit ecoa family putative	COG0732V	Defence mechanisms
SACOL0521	-	Hypothetical protein	COG0718S	Function unknown

ID	Gene name	Product	COG	Functional group
SACOL0526	-	DNA polymerase III delta prime subunit putative	COG0470L	DNA replication, recombination and repair
SACOL0733	-	Sugar efflux transporter putative	COG2814G	Carbohydrate transport and metabolism
SACOL0794	-	Hypothetical protein	Unknown	Unknown
SACOL0888	-	Pathogenicity island lipoprotein putative	Unknown	Unknown
SACOL1225	-	Hypothetical protein	Unknown	Unknown
SACOL1324	-	Hfq protein putative	COG1923R	General function prediction only
SACOL1739	-	Sensory box histidine kinase PhoR	COG5002T	Signal transduction mechanisms
SACOL2202	-	Hypothetical protein	Unknown	Unknown
SACOL2204	-	Hypothetical protein	Unknown	Unknown
SACOL2456	-	Hypothetical protein	COG4430S	Function unknown
SACOL2729	-	Integrase/recombinase core domain family	COG2826L	DNA replication, recombination and repair

Table 11. Genes displaying evidence for positive selection, with an elevated ratio of non-synonymous to synonymous mutations in non-human animal associated lineages when compared with human associated lineages, indicating positive selection in these lineages only

ID	Gene name	Product	COG	Functional group
SACOL0230	-	PTS system sorbitol-specific IIB component	COG3414G	Carbohydrate transport and metabolism
SACOL0414	-	Lipoprotein putative	COG2822P	Inorganic ion transport and metabolism
SACOL0117	-	Polysaccharide extrusion protein	COG2244R	General function prediction only
SACOL1384	*opuD1*	Osmoprotectant transporter BCCT family	COG1292M	Cell envelope biogenesis, outer membrane
SACOL2535	-	2-hydroxyacid dehydrogenase	COG1052CHR	Energy production and conversion, Coenzyme metabolism
SACOL1790	*murC*	UDP-N-acetylmuramate--L-alanine ligase	COG0773M	Cell envelope biogenesis, outer membrane
SACOL1294	-	Metallo-beta-lactamase family protein	COG0595R	General function prediction only
SACOL1719	*hemA*	Glutamyl-trna reductase	COG0373H	Coenzyme metabolism
SACOL0726	*sarX*	Staphylococcal accessory protein X	Unknown	Unknown

Table 12. Genes displaying evidence for positive selection with an elevated ratio of non-synonymous to synonymous mutations in human associated lineages when compared with other animal associated lineages, indicating positive selection in these lineages only.

ID	Gene name	Product	COG	Functional group
SACOL2174	-	Hypothetical protein	COG5547S	Function unknown
SACOL1270	*hslV*	ATP-dependent protease peptidase subunit	COG5405O	Posttranslational modification, protein turnover, chaperones
SACOL1277	*hslV*	ATP-dependent protease peptidase subunit	COG5405O	Posttranslational modification, protein turnover, chaperones
SACOL0960	*rocD*	Ornithine--oxo-acid transaminase	COG4992E	Amino acid transport and metabolism
SACOL1032	-	Comk family protein	COG4903K	Transcription
SACOL0938	*dltD*	Dltd protein	COG3966M	Cell envelope biogenesis, outer membrane
SACOL1378	-	Hypothetical protein	COG3763S	Function unknown
SACOL1393	-	Transcriptional antiterminator lict putative	COG3711K	Transcription
SACOL2003	*hlb*	B haemolysin	COG3568R	General function prediction only
SACOL0444	-	Hypothetical protein	COG3212S	Function unknown
SACOL0069	-	Hypothetical protein	COG3070K	Transcription
SACOL1151	-	Hypothetical protein	COG3027S	Function unknown
SACOL2034	-	Hypothetical protein	COG2391R	General function prediction only

ID	Gene name	Product	COG	Functional group
SACOL1937	*pepS*	Aminopeptidase peps	COG2309E	Amino acid transport and metabolism
SACOL1814	-	Lysophospholipase putative	COG2267I	Lipid metabolism
SACOL2057	*rsbU*	Sigma factor B regulator protein	COG2208TK	Signal transduction mechanisms, transcription
SACOL0806	*pepT*	Peptidase T	COG2195E	Amino acid transport and metabolism
SACOL1927	-	Hypothetical protein	COG1988R	General function prediction only
SACOL1681	-	Rrf2 family protein	COG1959K	Transcription
SACOL0538	*ispE*	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	COG1947I	Lipid metabolism
SACOL0096	*sarS*	Staphylococcal accessory regulator S	COG1846K	Transcription
SACOL0811	-	Hypothetical protein	COG1739S	Function unknown
SACOL2334	-	Hypothetical protein	COG1668CP	Energy production and conversion, inorganic ion transport and metabolism
SACOL2030	*scrR*	Sucrose operon repressor	COG1609K	Transcription
SACOL2577	*crtM*	Dehydrosqualene synthase	COG1562I	Lipid metabolism
SACOL0812	-	Degv family protein	COG1307S	Function unknown
SACOL2341	*fni*	Isopentenyl pyrophosphate isomerase	COG1304C	Energy production and conversion
SACOL1548	-	Atsa/elac family protein	COG1234R	General function prediction only

ID	Gene name	Product	COG	Functional group
SACOL2323	*hutI*	Imidazolonepropionase	COG1228Q	Secondary metabolites biosynthesis, transport and catabolism
SACOL2474	-	Peptide ABC transporter permease protein	COG1173EP	Amino acid transport and metabolism Inorganic ion transport and metabolism
SACOL0882	-	ABC transporter ATP-binding protein	COG1135P	Inorganic ion transport and metabolism
SACOL1613	-	ABC transporter ATP-binding protein	COG1121P	Inorganic ion transport and metabolism
SACOL2441	-	Amino acid permease	COG1113E	Amino acid transport and metabolism
SACOL1102	*pdhA*	Pyruvate dehydrogenase complex E1 component alpha subunit	COG1071C	Energy production and conversion
SACOL1984	*aldA2*	Aldehyde dehydrogenase	COG1012C	Energy production and conversion
SACOL1569	*nusB*	Transcription antitermination protein nusb	COG0781K	Transcription
SACOL0357	*dut*	Prophage L54a deoxyuridine 5'-triphosphate nucleotidohydrolase	COG0756F	Nucleotide transport and metabolism
SACOL0187	-	RGD-containing lipoprotein	COG0747E	Amino acid transport and metabolism
SACOL1420	-	Phosphate transport system protein phou putative	COG0704P	Inorganic ion transport and metabolism
SACOL2549	-	Esterase putative	COG0657I	Lipid metabolism

ID	Gene name	Product	COG	Functional group
SACOL2010	-	Iron compound ABC transporter iron compound-binding protein	COG0614P	Inorganic ion transport and metabolism
SACOL2277	-	Iron compound ABC transporter iron compound-binding protein	COG0614P	Inorganic ion transport and metabolism
SACOL1753	-	Universal stress protein family	COG0589T	Signal transduction mechanisms
SACOL2357	-	ABC transporter permease protein	COG0577V	Defence mechanisms
SACOL0823	*uvrB*	Excinuclease ABC subunit B	COG0556L	DNA replication, recombination and repair
SACOL1516	*rpsA*	Ribosomal protein S1	COG0539J	Translation, ribosomal structure and biogenesis
SACOL1935	-	Hypothetical protein	COG0535R	General function prediction only
SACOL1315	*hexA*	DNA mismatch repair protein	COG0249L	DNA replication, recombination and repair
SACOL2550	-	Hypothetical protein	COG0526OC	Energy production and conversion
SACOL1217	*pyrE*	Orotate phosphoribosyltransferase	COG0461F	Nucleotide transport and metabolism
SACOL1137	*rpmF*	50S ribosomal protein L32	COG0333J	Translation, ribosomal structure and biogenesis
SACOL0586	*rplL*	50S ribosomal protein L7/L12	COG0222J	Translation, ribosomal structure and biogenesis
SACOL2213	*rpoA*	DNA-directed RNA polymerase alpha subunit	COG0202K	Transcription
SACOL1229	*sun*	Sun protein	COG0144J	Translation, ribosomal structure and biogenesis
SACOL1566	*ispA*	Geranyltranstransferase	COG0142H	Coenzyme metabolism

ID	Gene name	Product	COG	Functional group
SACOL0018	*purA*	Adenylosuccinate synthetase	COG0104F	Nucleotide transport and metabolism
SACOL2215	*rpsM*	30S ribosomal protein S13	COG0099J	Translation, ribosomal structure and biogenesis
SACOL1792	-	Hypothetical protein	COG0073R	General function prediction only
SACOL2442	-	Na+/H+ antiporter putative	COG0025P	Inorganic ion transport and metabolism
SACOL0300	-	Hypothetical protein	Unknown	Unknown
SACOL0500	-	Hypothetical protein	Unknown	Unknown
SACOL0649	-	Hypothetical protein	Unknown	Unknown
SACOL0803	-	Lipoprotein putative	Unknown	Unknown
SACOL1980	-	Hypothetical protein	Unknown	Unknown
SACOL2672	-	Accessory secretory protein Asp3 putative	Unknown	Unknown
SACOL2690	*icaD*	Intercellular adhesion protein D	Unknown	Unknown
SACOL0788	-	Proton-dependent oligopeptide transporter family protein	Unknown	Unknown

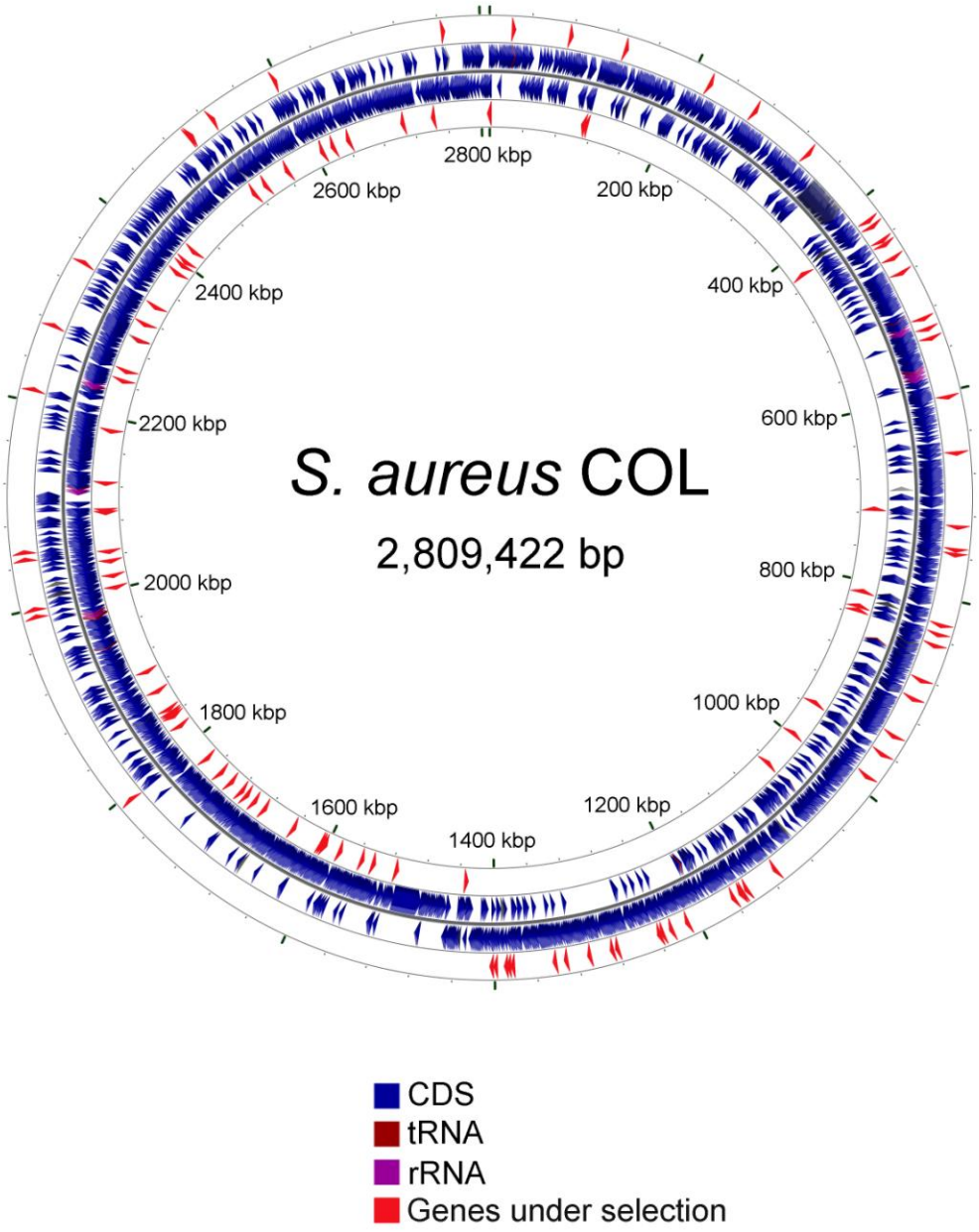


Figure 15. Positive selection acts upon genes distributed throughout the *S. aureus* chromosome. Genome sequence of *S. aureus* strain COL displaying coding sequences, tRNA and rRNA (blue, burgundy and purple arrows, respectively) and duplicates of genes under positive selection (red arrows). Inner circle displays size in kbp.

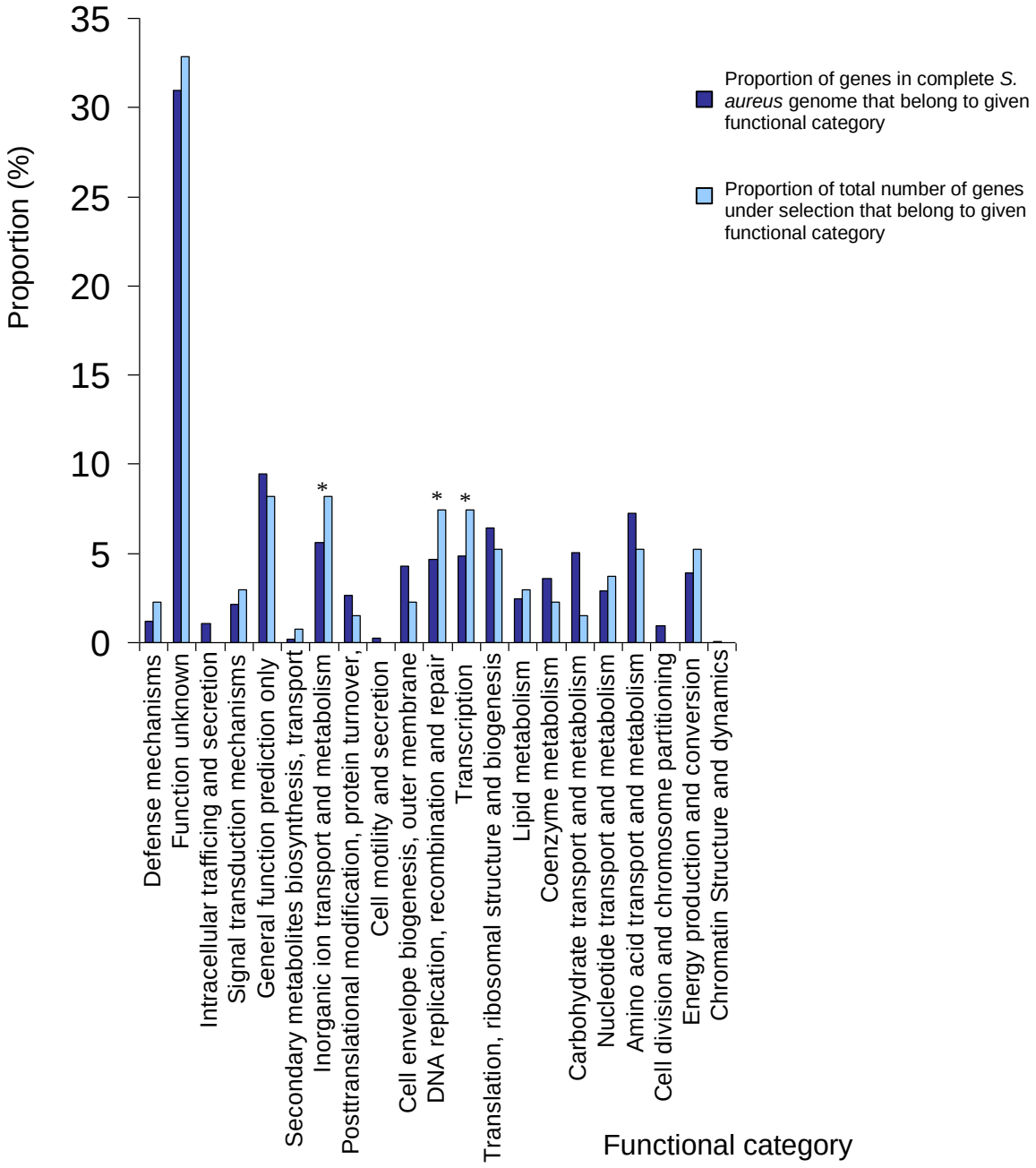


Figure 16. Functional categories 'inorganic ion/nucleotide transport', 'DNA replication/recombination/repair', and 'transcription and energy production/conversion' may be enriched with genes under positive selection in the *S. aureus* genome. COG functional categories are shown on the X axis and proprotion of genes in each category on the Y axis. Dark blue bars indicate proportion of genes in entire *S. aureus* COL genome that belong to given functional category; light blue bars indicate proportion of total number of genes under selection that belong to given functional category.

* p<0.1; COG category significantly enriched in positively selected genes.

5.4.3. Membrane transporters, gene regulators and DNA replication, recombination and repair proteins may be important for adaptation of *S. aureus* to diverse environments

In order to detect enrichment of particular functional categories with genes under positive selection, the proportion of genes in each category was compared between (i) all *S. aureus* genes, and (ii) genes under selection (Fig. 16). Genes under positive selection are slightly significantly more likely to encode proteins involved in DNA replication, recombination and repair, inorganic ion transport and metabolism, or transcription (χ^2 test, $p<0.1$; Fig. 16). However, it should be noted that a low probability such as this may also indicate that, although such proteins have probably been subject to selection, no particular category is under selection to a greater extent than any other.

Membrane transporter genes under selection encode many predicted ABC transporter components such as ATP-binding proteins, permease proteins and compound-binding proteins (Tables 10-12), and characterised transporter constituents such as osmoprotectant transporter OpuD1, Heme ABC transporters IsdD/F and iron compound ABC transporter sirA (Tables 10-12).

Proteins involved in transcriptional regulation are also under positive selection. These include the staphylococcal accessory regulators SarS and SarX, the sucrose operon repressor ScrR, a sigma factor B regulator protein RsbU and several genes with a putative role in transcriptional regulation (Tables 10-12). In addition, several genes under positive selection are sensory components of two component regulators (SACOL2646 and SACOL1739).

Proteins important for DNA replication, recombination and repair are also under selection. These include DNA mismatch repair protein HexA exonuclease SbcC involved with DNA repair, topoisomerase enzyme subunit ParC, and several others (Tables 10-12).

5.4.4. Known virulence genes are subject to positive selection

In order to determine the biological significance of genes under selection in terms of *S. aureus* pathogenicity, results were manually inspected to identify known virulence factors. These include accessory secretory protein Asp3 (part of a pathway resulting in expression of sraP), lipoproteins, an intercellular adhesion protein involved in biofilm formation (IcaD), iron compound-binding protein SirA and β toxin, a toxin with sphingomyelinase activity (Tables 10-12).

5.4.5. Proteins under selection in host-specific lineages

In order to ascertain whether positive selection is an important component of host adaptation by *S. aureus*, all genes under positive selection were analysed for evidence of selection acting upon a specific clonal lineage, indicated by elevated ω in one group only.

Rates in human-associated and non-human animal-associated clonal lineages were compared. In total, 9 proteins were found to be subject to positive selection in one or more strains associated with an animal host, including ED133 (ovine), RF122 (bovine), ED98 (avian) and DL190 (rabbit) but not with any strains of human origin. Tests did not further differentiate between different animal host species as clonal lineages were only represented by single strains, restricting the power of the test to identify host-specific positive selection. Elevated ratios were identified in several characterised proteins including Staphylococcal accessory protein X (SarX), Glutamyl-tRNA reductase (hemA), UDP-N-acetylmuramate-L-alanine ligase (MurC) and an osmoprotectant transporter of the BCCT family (OpuD1) (Table 11). Proteins with a general function assigned included a PTS system sorbitol-specific IIB component, a putative lipoprotein, a metallo-beta-lactamase family protein and a 2-hydroxyacid dehydrogenase (Table 11).

A total of 67 genes displayed an elevated ratio in human associated lineages only, encoding proteins from multiple functional categories including DNA replication, recombination and repair, inorganic ion/nucleotide/amino acid transport and

metabolism, translation and ribosomal structure and biogenesis, in addition to 9 hypothetical proteins (Table 12). Clonal lineages include major worldwide hospital associated clones CC5 and CC1, the CC30 and CC22 clones prevalent in UK hospitals, the CC8 clone responsible for a large proportion of community-acquired infections in the US, in addition to representatives of the CC7 (ST1159) and CC8 (ST72 and ST239) lineages isolated from healthy individuals. Seven genes had orthologues in human-associated lineages only, and must therefore be under selection in this group only.

A total of 44 genes were under selection in all lineages, indicating that the encoded proteins are not involved in host-specific evolution but instead have undergone change in response to universal selective pressures, or to pressures that relate to niche variation not related to host type.

5.5. Discussion

5.5.1. Positive selection in *S. aureus*

Evolution within the *S. aureus* species involves adaptation to a variety of ecological niches. These include different host species with distinctly structured immune systems and variation in host molecules with which bacteria interact, and host tissues that represent unique barriers to pathogen invasion. In addition, healthcare environments exert selective pressure for traits such as antimicrobial resistance, or altered transmission mechanisms via inert surfaces or healthcare workers. Comparative analysis of 30 *S. aureus* isolates from humans and other animals, and from diverse phylogenetic lineages, reveals genome-wide positive selection that has led to modification of proteins from the majority of functional categories, leading to diversification of the proteome, metabolome and replication capacity that may be associated with adaptation of *S. aureus* to diverse environments. These data imply that adaptation requires change to a broad range of cellular processes and biological traits.

Analysis of *S. aureus* genes under positive selection for enrichment in particular functional categories did not reveal very strong evidence that a particular category is under selection to a greater extent than any other. Nonetheless, a large number of genes under selection encode proteins involved in transport of inorganic ions or nucleotides. Adaptation to different milieu that vary in nutritional availability may require changes in the uptake of solutes across the membrane. For example, different lineages may specialise in infection of different host tissues, ranging from the human skin, blood or urogenital tract to the ruminant udder. Furthermore, as *S. aureus* is a commensal of the anterior nares and skin but also an opportunistic invader of other tissues, there is temporal variation in nutrient availability (Horsburgh, 2008). Such variation provides selective pressure for diversification of transporters involved in nutrient acquisition.

Of note, several of the transfer proteins found to be under positive selective pressure are involved in iron sequestration. Iron is a major determinant of bacterial survival (Ratledge & Dover, 2000) and is utilised by all bacterial pathogens due to its redox potential, important for an array of catalytic reactions (Miller & Britigan, 1997). Mammalian and avian iron is not freely available, instead being sequestered in complexes such as transferrins, ferritin, hemoglobin and other iron-containing proteins (Ratledge & Dover, 2000). Furthermore, the host further reduces the amount of available iron on detection of an invading pathogen in a process called hypoferraemia (Horsburgh, 2008). *S. aureus* has evolved several systems for releasing iron from the host environment (Fig. 17), a trait that enhances its success as a commensal and as a pathogen (Horsburgh, 2008).

One such system utilises the Isd family of proteins to sequester iron from haem via a complex, multicomponent pathway. Cell wall-anchored proteins carry haem (liberated from haemoglobin or haptoglobin) through the cell wall to a membrane associated

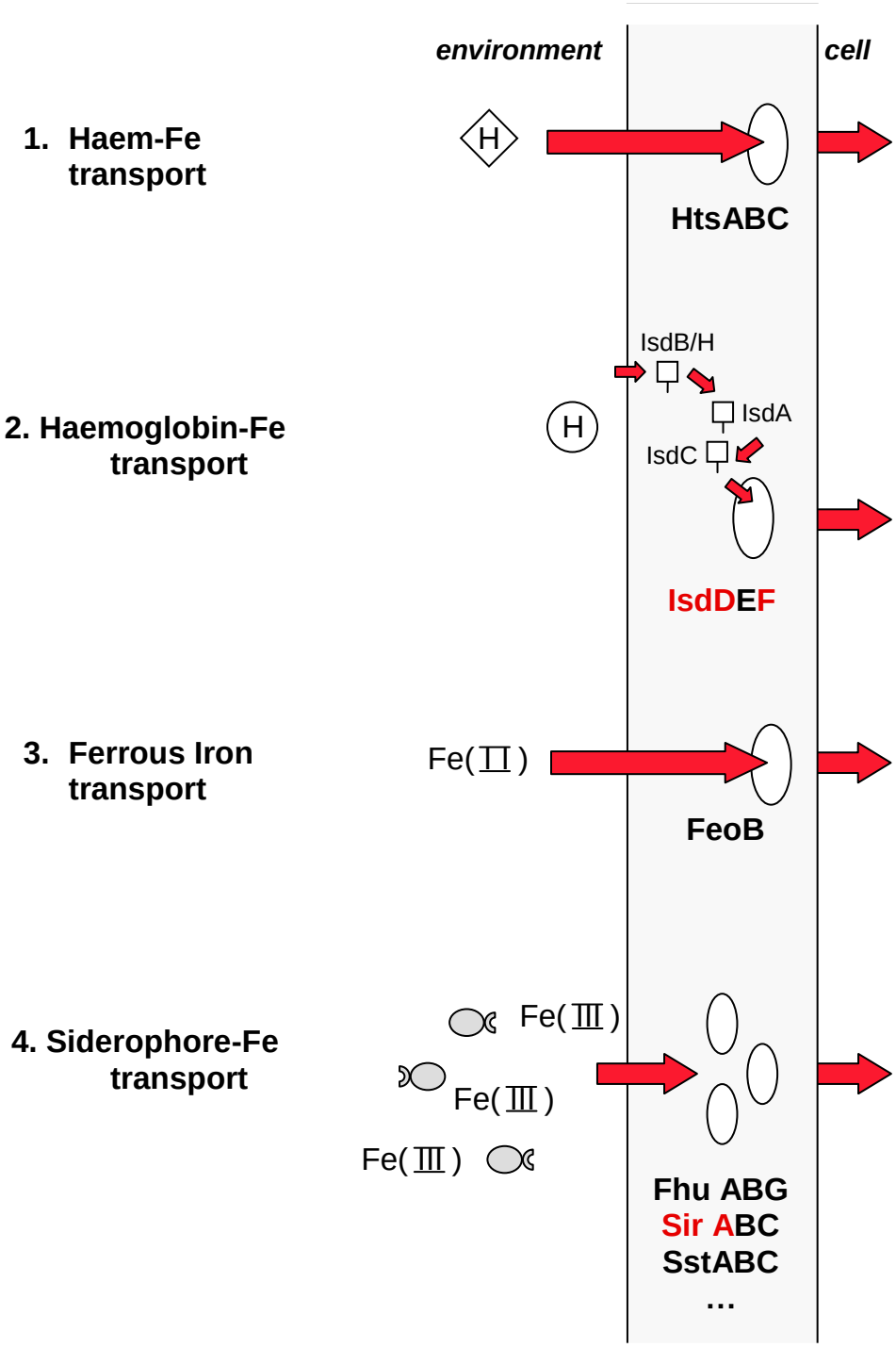


Figure 17. Genes under positive selection are components of different Fur-regulated mechanisms, employed by *S. aureus* for transport of iron into the cell. Diagram depicts the four major mechanisms of Fur-dependent iron sequestration and transport. Host iron sources include haem, haemoglobin, ferrous iron and iron sequestered via siderophore production. Siderophore-iron mediated transport involves several separate transport systems. Transport system in red are those where genes show evidence for positive selection (adapted from Horsburgh, 2008).

permease complex, which in turn transfers the complex to the cytoplasm where the tetrapole ring is cleaved to liberate iron (Skaar & Schneewind, 2004; Fig. 17). The genes encoding IsdD and IsdF, components of the membrane associated complex, show evidence for positive selection. This suggests that modification of the Isd pathway, and in particular the mechanism used to transport haem across the membrane, has been important for *S. aureus* evolution, which may promote survival capacity or allow increased pathogenicity in a particular setting.

A separate system involves production of siderophores, chelators of ferric iron that are secreted from the bacterium where they solubilise iron from a variety of sources (Ratledge & Dover, 2000). This is transported back into the microbial cell via a specific uptake carrier system, where the iron can be released following its reduction to ferrous ion (Ratledge & Dover, 2000). A number of separate systems exist for siderophore-mediated transport, one being SirABC (Dale *et al.*, 2004; Fig. 17), of which SirA was found to be under selection in the current study. As SirA is the binding component (lipoprotein) of this transporter system (Dale *et al.*, 2004) this may suggest that co-evolutionary adaptive change of the protein has occurred due to variation in iron compounds between different host environments.

In addition, four genes encoding uncharacterised iron compound ABC transporter components are under selection, including three putative permease proteins (SACOL1144, SACOL2165 and SACOL0705) and an iron compound-binding protein (SACOL2277). Taken together, these data indicate that acquisition of iron via a range of different mechanisms is an important process for *S. aureus*, and that adaptive evolution has involved change in multiple components of iron sequestration systems. This may be a response to variation in provision of iron-containing complexes in different niches or an adaptive strategy employed by the species in order to enhance overall retrieval of iron, an element essential for survival and pathogenicity (Ratledge & Dover, 2000).

In addition to the import of nutrients, ABC transporters in bacteria are responsible for active efflux of molecules (Davidson *et al.*, 2008). This can be important for transporting hydrophobic drugs from the cell or excreting lipids, peptides and proteins including toxins such as haemolysins (Davidson *et al.*, 2008). Thus, transporters may be adapting to the selective pressure of antibiotic presence within host tissues, or antimicrobial solutions utilised in the healthcare environment. Alternatively, diversification in transporters may affect secretion of bacterial virulence factors, which may in turn facilitate invasion of different environments. For example, selection may favour increased presence of exoproteins that inactivate host defensive molecules or of enzymes targeting host tissue components.

A second functional category of proteins with a large number of genes subject to positive selection is transcription, suggesting that selective pressures have led to change in regulatory genes. Change in regulators may affect expression of a wide range of genes, which can have a profound influence on the bacterium. For example, SarS and SarX (both of which are under selection) interact with the major two-component signal transduction system *agr* (Cheung *et al.*, 2001, Manna & Cheung, 2006) which is important for regulation of virulence factors (Yarwood & Schlievert, 2003). SarS is known to activate synthesis of SpA (Cheung *et al.*, 2001) and SarX is involved in expression of haemolysins (α toxin and β toxin) and proteases (a V8 protease, a zinc metalloprotease and a cysteine protease) (Manna & Cheung, 2006) Thus, adaptation via change in these regulators may increase or decrease pathogenicity, for example by altering the stage within bacterial growth at which virulence genes are expressed. Similarly, MecRI, an inducer of methicillin resistance protein MecA (Hiramatsu *et al.*, 1992), shows evidence for positive selection suggesting adaptation to an environment in which antimicrobial therapeutics are routinely utilised.

Regulation of genes involved in adaptation to different metabolic and nutritional requirements may also be affected. *S. aureus* has the capacity to respond to a variety of changes in environment, including stressors such as increase in temperature or salt or an

alkaline pH (Horsburgh, 2008). Many genes and systems are involved in this process but *rsbU*, encoding the sigma factor B regulator protein, is a key component of the *rsbUrsbVrsbWsigB* operon transcribed by σ^A (Kullik & Giachino, 1997, Senn *et al.*, 2005). RsbU functions as a key stress reception molecule (Kullik & Giachino, 1997, Senn *et al.*, 2005) and displays evidence for positive selection. In addition, expression of genes involved in metabolic processes will be influenced by SarX and SarS via regulation of *agr* (Cheung *et al.*, 2001, Dunman *et al.*, 2001, Manna & Cheung, 2006). The finding that such genes are subject to positive selection suggests that change in regulators may be necessary for adaptation to the different metabolic requirements of different niches.

A third functional category enriched with genes under positive selection is DNA replication, recombination and repair. Modification of these proteins may alter traits such as replication or mutation rate. Change in proteins such as the DNA mismatch repair protein HexA, whose homologues in other species are involved in hypermutation and recombination (Ginetti *et al.*, 1996, Humbert *et al.*, 1995, Mérino *et al.*, 2002, Prudhomme *et al.*, 1991), and topoisomerase enzyme subunit ParC (also known as GrlA), in which SNPs have been associated with antibiotic resistance (Hooper, 2002), may be important in a hospital environment where there is strong pressure for rapid acquisition of antibiotic resistance. An increase in mutation rate or in the number of generations in a given time period may increase the likelihood that mutations conferring resistance will arise, providing such clones with a selective advantage and a greater opportunity to disseminate. Conversely, in some settings selection will favour bacteria with a relatively low mutation rate as this reduces the likelihood of accumulation of detrimental mutations. Therefore, variation in selective pressures may lead to diversification in replication capacity across the species.

Many pathogenic phenotypes in *S. aureus* are conferred by proteins from a variety of functional categories, such as ligand-binding surface proteins, secreted enzymes and proteins required for immune evasion (Arvidson, 2006, Bohach, 2006, Foster, 2005,

Foster & Hook, 1998), including many that are under positive selection. For example, positive selection has led to diversification of the intercellular adhesion protein IcaD involved in biofilm formation (Gotz, 2002), Asp3, a component of a pathway which affects expression of the adhesin SraP (Siboo *et al.*, 2008), and a putative cell wall surface anchor family protein. Modification of such proteins may be important for colonisation of indwelling medical devices, which could enhance ability to cause nosocomial infections. Such infections are a key concern to healthcare providers and a major economic burden on society (Emmerson *et al.*, 1996, Jones, 2003). Adherence to components of the host extracellular matrix is also important as it mediates colonisation of host tissue.

The major staphylococcal virulence factor β toxin and a putative lyspohopholipase also display evidence for positive selection, suggesting that diversification of toxins confers a selective advantage. This may relate to a change in pathogenic potential or to variation in target molecules in different host species. Several lipoproteins, also thought to be involved in host-pathogen interaction and virulence, display evidence for positive selection. Furthermore, as discussed previously, the SirA, IsdD and IsdF proteins are involved in pathogenicity via their involvement in iron sequestration, and regulators SarX, SarS and RsbU play an important role in regulation of virulence gene expression.

These findings suggest that change in virulence factors may play a role in the evolution of *S. aureus*. Modification of proteins involved in colonisation or in provocation of disease symptoms may represent enhancement of virulence in *S. aureus*, evolving as a response to the selective pressures of the hospital environment. Bacteria may evolve greater virulence because more severe symptoms in the host lead to hospitalisation, providing bacteria with the opportunity to transmit to multiple immuno-compromised individuals via healthcare workers. Such proteins may promote host production of infectious materials such as pus and infected blood (Massey *et al.*, 2006). In addition, it has been suggested that a complex transmission pathway is a key factor in explaining the evolution and maintenance of virulence in *S. aureus*, in contrast to a species such as *S.*

epidermidis in which transmission is a single step process (Massey *et al.*, 2006). This ecological characteristic may provide a selective pressure towards increasingly pathogenic traits, as increased virulence allows *S. aureus* to overcome a complex transmission pathway.

Immune selection may play a crucial role in selecting for adaptive change in surface expressed proteins. The ability of the host immune system to recognise invading pathogens exerts a selective pressure favouring bacteria with variation in their surface molecules, as these may not be recognised and will therefore out-compete other clones, initiating an arms race between host and pathogen (Woolhouse *et al.*, 2002). Host adaptation may be a part of this process as structure and physiology of the immune system varies between different host taxa, necessitating variation in bacterial proteins with which the immune system may interact. In support of such a hypothesis, surface exposed proteins including putative lipoproteins and a putative cell wall surface anchor family protein were found to be under selection in the current study.

Of note, a large number of genes identified in this study encode proteins with no predicted function (designated hypothetical proteins). This finding suggests that many proteins that may be involved in virulence, niche adaptation or persistence in a particular environment are yet to be characterised. Such proteins represent ideal targets for future study to identify novel virulence determinants, enhancing our understanding of *S. aureus* pathogencicity and providing candidate targets for the development of novel therapeutics.

5.5.2. Host adaptation

In 9 genes, ratios were elevated in non-human animal-associated branches of the *S. aureus* phylogenetic tree, indicating positive selection acting on these strains only. The diversity of encoded protein functional categories reflects that observed across the species, in that proteins are involved in carbohydrate and inorganic ion transport and metabolism, transcription, coenzyme metabolism, cell envelope biogenesis and energy

production and conversion. This suggests that adaptation to a novel host type requires wide scale change to cellular processes and bacterial traits as opposed to, for example, change only in surface proteins that come into direct contact with the host and may thus be expected to be under immune selection.

For example, SACOL0230 encodes a PTS system sorbitol-specific IIB component, part of a major carbohydrate transport system in several species of bacteria (Postma *et al.*, 1993). It is likely that the carbohydrate or sugar content within the lumen of the bovine or ovine udder varies markedly from that of the human nasopharynx. Following a host switch, such a difference may strongly select for bacteria with altered metabolic activity that allows them to capitalise on differences in carbohydrate provision. Positive selection was also observed in OpuD1, a transporter of the b etaine-choline-carnitine (BCCT) family involved in osmoregulation in several species (Yang *et al.*, 2010, Kappes *et al.*, 1996, Lucht & Bremer, 1994). For example, the *opuD1* gene is up-regulated at high salt conditions in the moderately halophilic bacterium *Vibrio parahaemolyticus*, which is thought to lead to uptake and accumulation of solutes that provide osmoprotection (Yang *et al.*, 2010). This protein may be important for survival on the skin or mucosal membranes, which likely differ in salt content between host species.

Of note, the animal associated strains included in this analysis represent bovine and ovine lineages (ST151 and ST133, respectively) with a long-term animal association (Guinane *et al.*, 2010, Smyth *et al.*, 2009) in addition to avian and rabbit isolates (of ST5 and ST121 lineages, respectively), both of which are predicted to be the result of a recent human-to-animal host jump (Smyth *et al.*, 2009, this study). However, it is likely that the genes identified in this study represent those involved in long term host adaptation by ovine and bovine associated lineages. Positive selection leading to fixation of beneficial mutations is a relatively long term mechanism of genome evolution compared to HGT and gene loss that characterise avian adaptation by strains of the ST5 lineage. Thus, it is unlikely that sufficient time has elapsed since occurrence of human-to-avian/rabbit host jumps that ω has reached biological significance.

In 67 genes, non-synonymous mutations were more prevalent in human associated lineages than those of animal origin. This may indicate positive selection acting upon these genes in response to selective pressures unique to the human host niche. This may include invasion of human tissue which requires modification of extracellular enzymes, or interaction with the human immune system which may select for variation in surface proteins and those involved in immune evasion. Presumably, modification of these genes is not essential for niche adaptation in animal associated lineages. It is possible that the encoded proteins are not important in an animal host and thus are not involved in the adaptive process, consistent with the observation in this study that avian host adaptation involves the loss of function of putative human-specific proteins. Alternatively, it is possible that they represent the bias towards inclusion of human associated strains in the genome data set.

5.5.3. Future directions

Genome-wide scrutiny of the *S. aureus* genome for the molecular signature of positive selection reveals an array of genes which may be associated with adaptation by *S. aureus* to diverse environments. Structural modelling of encoded proteins, including location of residues under positive selection, may shed light on the molecular interaction between host and pathogen and the biological importance of particular protein domains or regions. Furthermore, functional characterisation and determination of the role of these genes, and the mutations that have arisen, is required in order to identify phenotypic changes that have been important for the evolution of *S. aureus*. Analysis of the complete genome including core, core-variable and accessory portions should also be considered in order that the full impact of positive selection upon genome evolution is realised.

Further scrutiny of different lineages would also provide great insight into the evolution of the species. Lineages that have a long term animal association but are unrelated to CC151, such as CC97 (which is primarily associated with a bovine host), were not

included in this study but could further clarify the host adaptive significance of the nine proteins identified in this study. In addition, inclusion of strains of the ST398 lineage would be of broad interest, as transmission of MRSA isolates in this lineage between pigs and humans appears to be prevalent (van Belkum *et al.*, 2008). Analysis of a larger number of animal isolates may also provide sufficient power to identify additional important proteins.

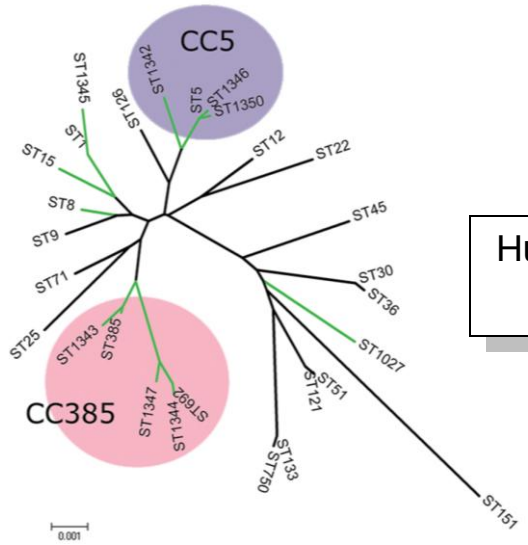
In addition, analysis of strains from human associated lineages currently under-represented in *S. aureus* genome sequence dataset may yield important results. Lineages that dominate strain collections from Africa and Asia are worthy of attention. Furthermore, prior epidemiological analysis may reveal unique selection pressures, enabling the development of novel hypotheses regarding niche adaptation.

The proteins identified in this study represent promising candidates for further study, in order to investigate the biological basis for host adaptation and host-specific pathogenicity by *S. aureus*. In addition to improving our understanding of pathogen evolution and host specificity, identification of proteins involved in the adaptive process may provide targets for therapeutics. *S. aureus* infection of livestock species represents a major welfare issue and economic burden on the farming industry, and novel targeted preventative or antimicrobial therapeutics are urgently required.

6. General Discussion

An understanding of pathogen ecology and evolution is required to reveal the molecular basis of infectious disease, and insights into host-pathogen dynamics and niche adaptation can prove valuable in the development of novel therapeutics and effective infection control strategies. In the past, knowledge of the biomedical basis of infection has prompted the development of antimicrobial agents, but recent years have seen a dramatic escalation in resistant microbes, demanding the urgent development of alternative approaches (Lebarbenchon *et al.*, 2008). An ultimate goal is to develop the ability to predict the conditions that cause parasites to evolve in a particular direction or to become more or less pathogenic (Lebarbenchon *et al.*, 2008), particularly if such conditions are anthropogenic. Human activities ranging from habitat and climate change to medical practice can have wide-ranging effects on microbial evolution, though their impact is often not appreciated. In addition, characterisation of the molecular mechanisms involved in this adaptive process and identification of essential genes and proteins are required to provide targets for novel therapeutics.

The current work reveals that the majority of *S. aureus* poultry isolates from farms in 6 countries on 4 continents belong to a sub-lineage of a major human associated clonal lineage, CC5. A single human-to-poultry host jump was followed by worldwide dissemination, presumably due to widespread transportation of live birds by the global poultry industry (Fig. 18). These data represent the first clearly defined example of a bacterial pathogen switching from a human to an animal host, within modern human history. A host jump from early humans to felines by *Helicobacter* occurred around 200,000 (range 50,000-400,000) years ago, and was followed by host adaptation leading to speciation (Eppinger *et al.*, 2006). It has been suggested that the jump occurred when a large feline ate a human infected with *H. pylori*, which led to colonisation of the animal by this human pathogen (Eppinger *et al.*, 2006). The diverged species *H. acinonychis* is now specific to large felines, and its genome displays evidence of host adaptation (Eppinger *et al.*, 2006). This event is a notable episode in the evolutionary history of the species, suggesting that exploitation of a novel niche can lead to species



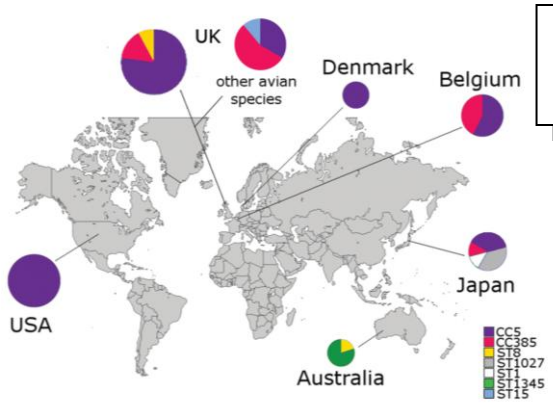
- Multi locus sequence typing
- 57 isolates from broiler chickens and other reared and wild birds

METHOD

Human to poultry host jump

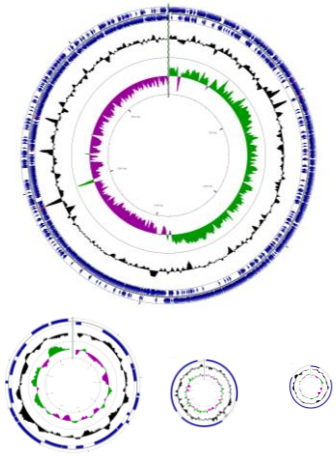
- Majority of poultry isolates CC5 lineage
- CC5 widespread and causes human disease
- Unrelated CC385 lineage
- Putative long term avian association

RESULTS



Worldwide dissemination

- 48 isolates from broiler chickens in 6 countries on 4 continents isolated over the last 54 years
- CC5 poultry clade wide geographic distribution
- Ancestral human CC5 geographic restriction
- Consistent with transmission via global poultry distribution network



- Whole genome sequencing of CC5 poultry and human isolates

METHOD

Adaptation to avian host

- Loss of function of proteins associated with human disease pathogenesis
- Acquisition of Mobile Genetic Elements from avian-specific accessory gene pool

RESULTS

Figure 18. Population genomic analysis indicates human-to-poultry host jump, widespread geographical dissemination and avian host-specific adaptation by *S. aureus.* Flow chart describes results of three stages of analysis performed in this study.

divergence in a bacterial pathogen, and is somewhat analogous to the host jump event in the CC5 poultry clade.

Much attention is paid to zoonotic pathogens and the emergence of animal-associated microbes in humans. However, the data presented here indicate that anthroponotic transfer may also be common. Of note, in addition to the CC5 clade, poultry strains belonging to several other human-associated lineages also show evidence for acquisition of MGE from an avian-specific accessory gene pool, suggesting that this phenomenon may not be limited to the CC5 clade. Anthroponotic transfer has important implications for the farming industry as human handlers may be a source of novel, emerging pathogens with the potential to become a major veterinary clinical problem. In light of the findings presented here, it is clear that further phylogenetic scrutiny of livestock pathogens is required in order to identify evidence for host switch events. It is advisable that regular screening and decontamination of flocks is implemented by the global poultry industry, and considered by other livestock industries.

In addition, further consideration should be given to the potential impact of globalised agriculture, which has drastically altered ecosystem dynamics and the nature of host-pathogen interaction for a large number of species. The poultry industry is a prime example of this. In the 21st Century the breeding, rearing and the import and export of live broiler chickens is a major global industry. However, prior to the 1920s, small-scale farmers reared chickens exclusively for their eggs, consuming chicken meat infrequently as a by-product. In post World War II America there was a conscious drive to improve breeding, disease control, nutrition, housing and processing within the industry (Boyd & Watts, 1997). Such improvements led to a rapid increase in production throughout the 1950s and 60s, along with vertical integration of the different stages of production to form large, national companies (Manning & Baines, 2004). Chicken production in the USA, as in many other developed countries, has grown steadily from the 1960s to the present day, rising from 2.6 million birds in 1961 to 22.3 million in 2006 (http://faostat.fao.org).

Growth was also observed in affiliated fields such as primary breeding, equipment supply, research and technical support from the pharmaceutical and chemical industries (Boyd & Watts, 1997). Breeding, in particular, has become a specialist field and a handful of multinational companies now supply a limited number of breeding lines to a global market. Such extensive change over a relatively short time period has resulted in a complex, multinational industry with few geographical barriers for the broiler chickens and little opportunity for establishment of localised ecosystems. Furthermore, intensive selective breeding of birds has resulted in a highly homogenous population, presumably reducing variation in immune response and susceptibility to infection by pathogens.

This study indicates that intercontinental transport of poultry by air has played a role in dissemination of an emerging poultry pathogen. Geographical distribution of the CC5 poultry clade contrasts strongly with its human-associated CC5 ancestors, and with phylogeographic clustering observed in other *S. aureus* lineages such as ST239, in which geographical restriction predominates (Nubel *et al.*, 2008, Harris *et al.*, 2010). The implications of this finding are substantial, as it suggests that human activity can have a major influence on the transmission dynamics of a livestock pathogen, and also that pandemic spread of a single clade of *S. aureus* is possible given appropriate conditions.

Globalisation and the rise of international air travel have also had a major impact on the spread of pathogens between humans. For example, rapid dissemination of the Severe Acute Respiratory Syndrome (SARS) virus in 2002-2003 was due at least in part to air travel (Knobler *et al.*, 2004). Patients suffering from SARS were initially restricted to China until the coronavirus responsible was transmitted along major airline routes to Hong Kong, Singapore, Hanoi and Toronto (Knobler *et al.*, 2004; Fig. 16). As the SARS outbreak became a major concern for healthcare providers worldwide, "an unprecedented multifaceted, multilateral, and multidisciplinary response" was coordinated by the WHO (Knobler *et al.*, 2004). The response to the SARS pandemic involved daily contact between researchers, diagnostic laboratories, clinicians, public

health institutions, ministries of health and WHO Country Offices (Knobler *et al.*, 2004). It is clear that such a large scale, coordinated response was critical to containing or at least minimising the pandemic, but such a contingency plan must be prepared in advance and be based upon sound scientific and clinical knowledge of the potential origin of an outbreak. The study presented here has highlighted such a potential origin, and should be considered by policy makers.

Of note, the only country among those sampled in which CC5 poultry strains were not identified was Australia. The Department of Agriculture, Fisheries and Forestry Biosecurity (Australia) impose very tight import/export rules in an attempt to minimize the introduction of alien species to the Australian ecosystem, including import of livestock animals (http://www.daff.gov.au/). Until relatively recently the import of poultry was banned entirely, and in the present day strict quarantine measures are imposed. If the farming industry and/or international legislators are to minimise the occurrence of pathogen pandemics, such as that identified in this study, it is perhaps advisable that similar restrictions are imposed elsewhere, in conjunction with screening programmes to identify emerging pathogens.

In addition to the ecological and epidemiological basis of a host switch, data presented here reveal a great deal about the underlying biology of host adaptation in *S. aureus* (Fig. 19).

This study has revealed the key mechanisms of genome diversification associated with host adaptive evolution by *S. aureus*. Loss of function of proteins implicated in human pathogenesis and acquisition of novel MGE from an avian-specific accessory gene pool have been observed in the CC5 clade. Furthermore, genome scale analysis of 30 strains isolated from human and animal hosts reveals evidence for positive selection, associated in some instances with host specificity. Taken together, these data suggest that host adaptation in *S. aureus* involves, over the short term, large scale genome change via HGT and smaller scale change to individual or a small number of nucleotides leading to

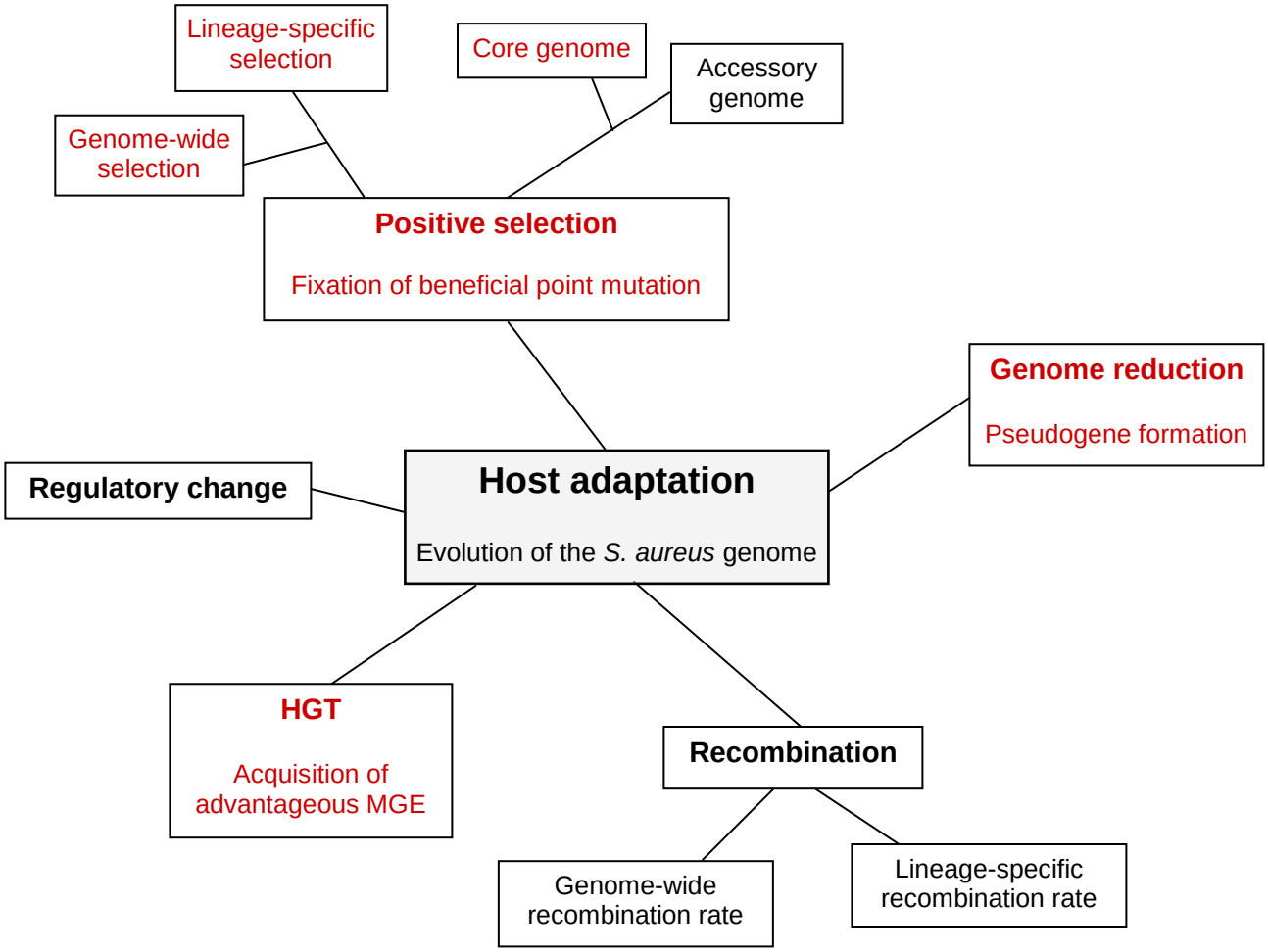


Figure 19. Molecular mechanisms of genome evolution in *S. aureus* include HGT of advantageous MGE, positive selection, genome reduction and recombination. Red text indicates areas considered in this study, black text indicates areas still to be investigated in relation to niche adaptation.

pseudogene formation. Over the longer term, positive selection acting upon the genome can lead to variation in amino acids which may affect protein structure.

The importance of HGT of MGE as a component of *S. aureus* evolution has been recognised in several studies (for review see Lindsay, 2008). These frequently focus on the acquisition of antibiotic resistance determinants such as SCC*mec* and the *vanA* transposon (Chang *et al.*, 2003, Katayama *et al.*, 2000, Weigel *et al.*, 2003), or on genome-wide evidence for HGT between lineages or species that shape the virulence and resistance profile of a given strain (An Diep *et al.*, 2006, Baba *et al.*, 2002, Gill *et al.*, 2005, Holden *et al.*, 2004). This study has revealed the importance of acquisition of MGE from a host-specific accessory gene pool, maintained in strains of diverse evolutionary origin but sharing an avian host niche. Though resistance and virulence determinants confer a major selective advantage for many *S. aureus* strains, this study also highlights the importance of MGE carrying genes with a putative role in metabolism, transport and interaction with variable host immune systems.

Recent population genomic analysis of *S. aureus* ruminant isolates provided evidence that the CC133 clade also evolved as the result of a human-to-ruminant host jump approximately 115–1,204 years ago followed by adaptive genome diversification (Guinane *et al.*, 2010). Of note, the genome of strain ED133 contains a unique complement of MGEs, including two new SaPIs and three novel prophages not previously identified in sequenced *S. aureus* strains, which are widely distributed in CC133 isolates (Guinane *et al.*, 2010). Genes on these MGE encode proteins putatively important for ruminant host adaptation, including ovine-specific variants of TSST-1 and SEC (encoded by SaPIov1) which have previously been shown to vary in biological activity in comparison with the alleles produced by human or bovine strains (Lee *et al.*, 1992, Deringer *et al.*, 1997). Further, phenotypic analysis revealed that the presence of SaPIov2, which encodes a variant of von Willebrand binding protein ($vWbp^{Sov2}$), confers the ability to coagulate ruminant plasma suggesting an important role in ruminant disease pathogenesis (Guinane *et al.*, 2010). Taken together, results of these

two studies suggest that acquisition of MGE is a major component of host adaptation of *S. aureus*.

Gene loss is recognised as a feature of adaptation to a novel and/or restricted niche in a variety of species, ranging from strains of *Helicobacter pylori* and *Salmonella enterica* that have undergone a host switch (Eppinger *et al.*, 2006, Thomson *et al.*, 2008) to the intracellular pathogen *Mycobacterium leprae* in which extreme genome reduction is observed (Cole *et al.*, 2001). Previous analyses have considered the prevalence of pseudogene formation across the *S. aureus* species as a whole, comparing the proportion of the genome in which genes are no longer intact between different species (Lerat & Ochman, 2005, Liu *et al.*, 2004). However, this is the first report of such a process playing an important role in adaptation in *S. aureus*.

Similarly, the importance of positive selection has been recognised in several species. These include uropathogenic lineages of *E. coli*, in which genes important for persistence and pathogenesis in the urinary tract environment were under selection (Chen *et al.*, 2006a), and *Campylobacter* in which 92.5 % of the nonrecombinant core genome is under positive selection in at least one lineage (Lefebure & Stanhope, 2009). However, this study represents the first genome-wide analysis of positive selection across diverse lineages of *S. aureus*. Data presented here indicate that small scale change to proteins throughout the core genome should be considered a significant source of adaptive variation in the species.

In order to evaluate the adaptive significance of such genomic change, *in vitro* assays could be performed to establish phenotypic differences between different strains. For example, variation in rate or manner of nutrient acquisition or immune response may be observed between CC5 poultry isolates and basal human associated strains, which can be attributed to a selective advantage for a given phenotype in infection of a particular host taxon. In a related study we have discovered that poultry strain ED98 has enhanced resistance to killing by avian heterophils compared with the closely related human MR1

strain (Lowder *et al.*, 2009). By contrast, there was no significant difference between the strains in resistance to phagocytic killing by human neutrophils. These data suggest that genetic differences identified between the ED98 genome and the human MR1 genome contribute to avian host-specific innate immune avoidance and reflect avian host adaptation.

It is likely that the avian-specific MGE identified here carry genes advantageous for colonisation or pathogenesis in an avian host. In order to establish the functional significance of proteins encoded by genes on these MGE, appropriate *in vitro* assays should be performed with isogenic strains deficient in a given MGE, followed by complementation by a plasmid encoding specific genes. In parallel, experimental infection of chickens with strains in which specific MGE have been deleted, in comparison to wild-type strains, may indicate attenuation in virulence or survival which would reflect a role for the MGE in host specific interaction. In the case of single nucleotide changes due to positive selection, it would be of interest to assess the structural significance of these changes by protein modelling, crystallisation and functional analysis.

In conclusion, this study provides a unique insight into host-adaptive evolution of *S. aureus*, and the impact of human activity on adaptation and dissemination of a major bacterial pathogen. In light of the findings presented here, it is important that the agricultural industry consider the impact of anthroponotic transfer of pathogens and act accordingly. Furthermore, the molecular genetic basis of host adaptation has been elucidated, providing a sound basis for future studies of pathogen evolution and the development of veterinary therapeutics and control strategies.

References

Achtman, M. (2008) Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annu Rev Microbiol* **62**: 53-70.

Aires-de-Sousa, M., Parente, C. E. S. R., Vieira-da-Motta, O., Bonna, I. C. F., Silva, D. A. & de Lencastre, H. (2007) Characterization of *Staphylococcus aureus* Isolates from Buffalo, Bovine, Ovine, and Caprine Milk Samples Collected in Rio de Janeiro State, Brazil. *Appl Environ Microbiol* **73**(12): 3845-3849.

An Diep, B., Gill, S. R., Chang, R. F., Phan, T. H., Chen, J. H., Davidson, M. G., Lin, F., Lin, J., Carleton, H. A., Mongodin, E. F., Sensabaugh, G. F. & Perdreau-Remington, F. (2006) Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *The Lancet* **367**(9512): 731-739.

Anderson, D. J. & Kaye, K. S. (2009) Surgical-Site Infections and Surgical Prophylaxis. In: *Staphylococci in Human Disease, 2nd edition.* K. B. Crossley, K. K. Jefferson, G. L. Archer & V. G. Fowler (eds). Oxford, UK: Wiley-Blackwell.

Anisimova, M. & Liberles, D. A. (2007) The quest for natural selection in the age of comparative genomics. *Heredity* **99**(6): 567-579.

Armand-Lefevre, L., Ruimy, R. & Andremont, A. (2005) Clonal Comparison of *Staphylococcus aureus* Isolates from Healthy Pig Farmers, Human Controls, and Pigs. *Emerg. Infect. Dis.* **11**(5): 711-714.

Arvidson, S. (2006) Extracellular Enzymes. In: *Gram-Positive Pathogens, 2nd edition.* V. A. Fischetti, R. Novick, J. Ferretti, D. Portnoy & J. Rood (eds). Washington, D.C.: ASM Press.

Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Meyer, F., Olsen, G., Olson, R., Osterman, A., Overbeek, R., McNeil, L., Paarmann, D., Paczian, T., Parrello, B., Pusch, G., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* **9**(1): 75.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. (2008) Genome Sequence of *Staphylococcus aureus* Strain Newman and Comparative Analysis

of Staphylococcal Genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol* **190**(1): 300-310.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K.-i., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. & Hiramatsu, K. (2002) Genome and virulence determinants of high virulence community-acquired MRSA. *The Lancet* **359**: 1819-1827.

Balbi, K. J., Rocha, E. P. C. & Feil, E. J. (2008) The temporal dynamics of slightly deleterious mutations in *Escherichia coli* and *Shigella* spp. *Mol Biol Evol* 26 (2): 345-355.

Barkas, T. & Watson, C. M. (1979) Induction of an Fc conformational change by binding of antigen: the generation of protein A-reactive sites in chicken immunoglobulin. *Immunology* **36**(3): 557-561.

Ben Zakour, N. L., Guinane, C. M. & Fitzgerald, J. R. (2008) Pathogenomics of the staphylococci: insights into niche adaptation and the emergence of new virulent strains. *FEMS Microbiol Lett* **289**(1): 1-12.

Bertolatti, D., O'Brien, F. G. & Grubb, W. B. (2003) Characterization of drug-resistant *Staphylococcus aureus* isolated from poultry processing plants in Western Australia. *Int J Env Health Res* **13**: 43-54.

Biggs, A. (2009) *Mastitis in Cattle*. Wiltshire, UK: The Crowood Press Ltd.

Bohach, G. (2006) *Staphylococcus aureus* Exotoxins. In: *Gram-Positive Pathogens, 2nd edition.* V. A. Fischetti, R. Novick, J. Ferretti, D. Portnoy & J. Rood (eds). Washington, D.C.: ASM Press.

Boyd, W. & Watts, M. (1997) Agro-Industrial Just-in-Time: The Chicken Industry and Postwar American Capitalism. In: *Globalising Food: Agrarian Questions and Global Restructuring.* D. Goodman & M. Watts (eds). London: Routledge.

Brown, N. F., Wickham, M. E., Coombes, B. K. & Finlay, B. B. (2006) Crossing the Line: Selection and Evolution of Virulence Traits. *PLoS Pathogens* **2**(5): 346-353.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422-3423.

Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., Shah, S., Rudrik, J. T., Pupp, G. R., Brown, W. J., Cardo, D., Fridkin, S. K. & T he Vancomycin-Resistant Staphylococcus aureus Investigative Team (2003) Infection with Vancomycin-Resistant *Staphylococcus aureus* Containing the vanA Resistance Gene. *N Engl J Med* **348**(14): 1342-1347.

Chen, H., Smith, G. J. D., Li, K. S., Wang, J., Fan, X. H., Rayner, J. M., Vijaykrishna, D., Zhang, J. X., Zhang, L. J., Guo, C. T., Cheung, C. L., Xu, K. M., Duan, L., Huang, K., Qin, K., Leung, Y. H. C., Wu, W. L., Lu, H. R., Chen, Y., Xia, N. S., Naipospos, T. S. P., Yuen, K. Y., Hassan, S. S., Bahri, S., Nguyen, T. D., Webster, R. G., Peiris, J. S. M. & Guan, Y. (2006a) Establishment of multiple sublineages of H5N1 influenza virus in Asia: Implications for pandemic control. *P Nat Acad Sci USA* **103**: 2845-2850.

Chen, S. L., Hung, C., Xu, J., Reigstad, C., Magrini, V., Sabo, A., Blasiar, D., Bieri, T., Meyer, R., Ozersky, P., Armstrong, J., Fulton, R., Latreille, J., Spieth, J., Hooton, T., Mardis, E., Hultgren, S. & Gordon, J. (2006b) Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli:* a comparative genomics approach. *P Nat Acad Sci USA* **103**(15): 5977-5982.

Cheung, A. L., Schmidt, K., Bateman, B. & Manna, A. C. (2001) SarS, a SarA Homolog Repressible by agr, Is an Activator of Protein A Synthesis in *Staphylococcus aureus*. *Infect. Immun.* **69**(4): 2448-2455.

Choi, Y. W., Kotzin, B., Herron, L., Callahan, J., Marrack, P. & Kappler, J. (1989) Interaction of *Staphylococcus aureus* toxin "superantigens" with human T cells. *P Nat Acad Sci USA* **86**(22): 8941-8945.

Clark, A. G., Glanowski, S., Nielsen, R., Thomas, P. D., Kejariwal, A., Todd, M. A., Tanenbaum, D. M., Civello, D., Lu, F., Murphy, B. *et al.* (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* **302**: 1960-1963.

Clark, N. C., Weigel, L. M., Patel, J. B. & Tenover, F. C. (2005) Comparison of Tn1546-Like Elements in Vancomycin-Resistant *Staphylococcus aureus* Isolates from Michigan and Pennsylvania. *Antimicrob Agents Ch* **49**(1): 470-472.

Clarke, S. R. & Foster, S. J. (2006) Surface adhesins of *Staphylococcus aureus*. *Adv Microb Physiol* **51**: 187-224.

Clutton-Brock, J. (1981) *Domesticated Animals From Early Times*. Austin, T.X., USA: Univ. Texas Press.

Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R. M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J. R. & Barrell, B. G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**(6823): 1007-1011.

Collen, D. (1998) Staphylokinase: a potent, uniquely fibrin-selective thrombolytic agent. *Nat Med* **4**(279).

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695-777.

Corey, G. Â R. (2009) *Staphylococcus aureus* Bloodstream Infections: Definitions and Treatment. *Clin Infect Dis* **48**(s4): S254-S259.

Corrigan, R. M., Rigby, D., Handley, P. & Foster, T. J. (2007) The role of *Staphylococcus aureus* surface protein SasG in adherence and biofilm formation. *Microbiology* **153**(8): 2435-2446.

Cosgrove, S. E., Carroll, K. C. & Perl, T. M. (2004) *Staphylococcus aureus* with Reduced Susceptibility to Vancomycin. *Clin Infect Dis* **39**(4): 539-545.

Cox, H. U. (2006) Staphylococcal infections. In: *Infectious Diseases of the Dog and Cat.* C. E. Greene (ed). Missouri, USA: Elsevier, Inc.

Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I. & Penades, J. R. (2001) Bap, a *Staphylococcus aureus* Surface Protein Involved in Biofilm Formation. *J Bacteriol* **183**(9): 2888-2896.

Dale, S. E., Sebulsky, M. T. & Heinrichs, D. E. (2004) Involvement of SirABC in Iron-Siderophore Import in *Staphylococcus aureus*. *J Bacteriol* **186**(24): 8356-8362.

Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. (2008) Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiol Mol Biol Rev* **72**(2): 317-364.

Davidson, L. E. & Boucher, H. W. (2009) Infections in Immunocompromised Patients. In: *Staphylococci in Human Disease, 2nd edition.* K. B. Crossley, K. K. Jefferson, G. L. Archer & V. G. Fowler (eds). Oxford, UK: Wiley-Blackwell.

Deringer, J.R., Ely, R.J., Monday, S.R., Stauffacher, C.V., Bohach, G.A. (1997) Vbeta-dependent stimulation of bovine and human T cells by hostspecific staphylococcal enterotoxins. *Infect Immun* 65: 4048–4054.

Devriese, L. A. (1984) A simplified system for biotyping *Staphylococcus aureus* strains isolated from different animal species. *J Appl Bacteriol* **56**: 215-220.

Devriese, L. A. & Oeding, P. (1976) Characteristics of *Staphylococcus aureus* strains isolated from different animal species. *Res. Vet. Sci.* **21**(3): 284-291.

Diamond, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature* **418**(6898): 700-707.

Dryden, M. S. (2009) Skin and soft tissue infection: microbiology and epidemiology. *Int J Antimicrob Ag* **34**(Supplement 1): S2-S7.

Dubin, G. (2002) Extracellular Proteases of Staphylococcus spp. *Biol Chem* **383**(7-8): 1075-1086.

Dunman, P. M., Murphy, E., Haney, S., Palacios, D., Tucker-Kellogg, G., Wu, S., Brown, E. L., Zagursky, R. J., Shlaes, D. & Projan, S. J. (2001) Transcription Profiling-Based Identification of *Staphylococcus aureus* Genes Regulated by the agr and/or sarA Loci. *J Bacteriol* **183**(24): 7341-7353.

Dussart, P., Cartet, G., Huguet, P., Lévêque, N., Hajjar, C., Morvan, J., Vanderkerckhove, J., Ferret, K., Lina, B., Chomel, J.-J. & Norder, H. (2005)

Outbreak of acute hemorrhagic conjunctivitis in French Guiana and West Indies caused by coxsackievirus A24 variant: Phylogenetic analysis reveals Asian import. *J Med Virol* **75**(4): 559-565.

Duthie, E. S. & Lorenz, L. L. (1952) Staphylococcal coagulase; mode of action and antigenicity. *J Gen Microbiol* **6**(95-107).

Emmerson, A., Enstone, J., Griffin, M., Kelsey, M. & Smyth, E. (1996) The Second National Prevalence Survey of infection in hospitals - overview of the results. *J Hosp Infect* **32:**(3): 175-190.

Enright, M. C., Day, N. P. J., Davies, C. E., Peacock, S. J. & Spratt, B. G. (2000) Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus*. *J Clin Microbiol* **38**(3): 1008-1015.

Enright, M. C., Robinson, D. A., Randle, G., Feil, E. J., Grundmann, H. & Spratt, B. G. (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *PNAS* **99**(11): 7687-7692.

Eppinger, M., Baar, C., Linz, B., Raddatz, G., uuml, nter, Lanz, C., Keller, H., Morelli, G., Gressmann, H., Achtman, M. & Schuster, S. C. (2006) Who ate whom? Adaptive Helicobacter genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genetics* (2006): e120.eor.

Ewing, B., Hillier, L., Wendl, M. & Green, P. (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175-185.

Falush, D., Wirth, T., Linz, B., Pritchard, J. K., Stephens, M., Kidd, M., Blaser, M. J., Graham, D. Y., Vacher, S., Perez-Perez, G. I., Yamaoka, Y., Megraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M. & Suerbaum, S. (2003) Traces of Human Migrations in *Helicobacter pylori* Populations. *Science* **299**(5612): 1582-1585.

Fan, J., Shu, M., Zhang, G., Zhou, W., Jiang, Y., Zhu, Y., Chen, G., Peacock, S. J., Wan, C., Pan, W. & Feil, E. J. (2009) Biogeography and Virulence of *Staphylococcus aureus*. *PLoS ONE* **4**(7): e6216.

Farrell, D. J., Mendes, R. E., Ross, J. E. & Jones, R. N. (2009) Linezolid surveillance program results for 2008 (LEADER Program for 2008). *Diagn Micr Infec Dis* **65**(4): 392-403.

Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., Moore, C. E. & Day, N. P. J. (2003) How clonal is *Staphylococcus aureus*? *J Bacteriol* **185**(11): 3307-3316.

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* **186**(5): 1518-1530.

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.

Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. (2001) The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science* **292**(5519): 1155-1160.

Fiennes, R. (1978) *Zoonoses and the Origins and Ecology of Human Disease*. London, UK: Academic Press.

Fitzgerald, J. R., Meaney, W. J., Hartigan, P. J., Smyth, C. J. & Kapur, V. (1997) Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. *Epidemiol Infect* **119**(2): 261-269.

Fitzgerald, J. R., Monday, S. R., Foster, T. J., Bohach, G. A., Hartigan, P. J., Meaney, W. J. & Smyth, C. J. (2001a) Characterization of a Putative Pathogenicity Island from Bovine *Staphylococcus aureus* Encoding Multiple Superantigens. *J Bacteriol* **183**(1): 63-70.

Fitzgerald, J. R., Reid, S. D., Ruotsalainen, E., Tripp, T. J., Liu, M., Cole, R., Kuusela, P., Schlievert, P. M., Jarvinen, A. & Musser, J. M. (2003) Genome Diversification in *Staphylococcus aureus*: Molecular Evolution of a Highly Variable Chromosomal Region Encoding the Staphylococcal Exotoxin-Like Family of Proteins. *Infect Immun* **71**(5): 2827-2838.

Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. (2001b) Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *P Nat Acad Sci USA* **98**(15): 8821-8826.

Foster, T. J. (2005) Immune evasion by staphylococci. *Nat Rev Micro* **3**(12): 948-958.

Foster, T. J. & Hook, M. (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol* **6**(12): 484-488.

Fournier, B. (2008) Global regulators of *Staphylococcus aureus* virulence genes. In: *Staphylococcus molecular genetics.* J. A. Lindsay (ed). Norfolk, UK: Caister Academic Press.

Fowler, V. G., Nelson, C. L., McIntyre, L. M., Kreiswirth, B. N., Monk, A., Archer, G. L., Federspiel, J., Naidich, S., Remortel, B., Rude, T., Brown, P., Reller, L. B., Corey, G. R. & Gill, S. R. (2007) Potential Associations between Hematogenous Complications and Bacterial Genotype in *Staphylococcus aureus* Infection. *J Infect Dis* **196**: 738-747.

Fowler, V. J., Miro, J., Hoen, B., Cabell, C., Abrutyn, E., Rubinstein, E., Corey, G., Spelman, D., Bradley, S., Barsic, B., Pappas, P., Anstrom, K., Wray, D., Fortes, C., Anguera, I., Athan, E., Jones, P., van der Meer, J., Elliott, T., Levine, D., Bayer, A. & ICE Investigators (2005) *Staphylococcus aureus* endocarditis: a consequence of medical progress. *JAMA* **293**(24): 3012-3021.

Gemmell, C., Tree, R., Patel, A., O'Reilly, M. & Foster, T. (1991) Susceptibility to opsonophagocytosis of protein A, α-haemolysin and β-toxin deficient mutants of *Staphylococcus aureus* isolated by allele-replacement. *J Zentralbl Bakteriol* **21**(273-277): Suppl.

George, L. W., Divers, T. J., Ducharme, N. & Welcome, F. L. (2008) Diseases of the Teats and Udder. In: *Rebhun's Diseases of Dairy Cattle.* T. J. Divers & S. F. Peek (eds). Missouri, USA: Elsevier Inc.

Gill, S. R., Fouts, D. E., Archer, G. L., Mongodin, E. F., DeBoy, R. T., Ravel, J., Paulsen, I. T., Kolonay, J. F., Brinkac, L., Beanan, M., Dodson, R. J., Daugherty, S. C., Madupu, R., Angiuoli, S. V., Durkin, A. S., Haft, D. H., Vamathevan, J.,

Khouri, H., Utterback, T., Lee, C., Dimitrov, G., Jiang, L., Qin, H., Weidman, J., Tran, K., Kang, K., Hance, I. R., Nelson, K. E. & Fraser, C. M. (2005) Insights on Evolution of Virulence and Resistance from the Complete Genome Analysis of an Early Methicillin-Resistant *Staphylococcus aureus* Strain and a Biofilm-Producing Methicillin-Resistant *Staphylococcus epidermidi*s Strain. *J Bacteriol* **187**(7): 2426-2438.

Gillaspy, A. F., Worrell, V., Orvis, J., Roe, B. A., Dyer, D. W. & Iandolo, J. J. (2006) *Staphylococcus aureus* NCTC8325 genome. In: *Gram-positive pathogens.* V. Fischetti, R. Novick, J. Ferretti, D. Portnoy & J. Rood (eds). Washington, DC, USA: ASM Press,.

Ginetti, F., Perego, M., Albertini, A. M. & Galizzi, A. (1996) *Bacillus subtilis* mutS mutL operon: identification, nucleotide sequence and mutagenesis. *Microbiology* **142**(8): 2021-2029.

Giuliani, M. M., Adu-Bobie, J., Comanducci, M., AricÃ, B., Savino, S., Santini, L., Brunelli, B., Bambini, S., Biolchi, A., Capecchi, B., Cartocci, E., Ciucchi, L., Di Marcello, F., Ferlicca, F., Galli, B., Luzzi, E., Masignani, V., Serruto, D., Veggi, D., Contorni, M., Morandi, M., Bartalesi, A., Cinotti, V., Mannucci, D., Titta, F., Ovidi, E., Welsch, J. A., Granoff, D., Rappuoli, R. & Pizza, M. (2006) A universal vaccine for serogroup B meningococcus. *P Nat Acad Sci USA* **103**(29): 10834-10839.

Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., Broker, B. M., Doskar, J. & Wolz, C. (2009) Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J Bacteriol* **191**(11).

Gomez, M. I., Lee, A., Reddy, B., Muir, A., Soong, G., Pitt, A., Cheung, A. & Prince, A. (2004) Staphylococcus aureus protein A induces airway epithelial inflammatory responses by activating TNFR1. *Nat Med* **10**(8): 842-848.

Gomez, M. I., O'Seaghdha, M., Magargee, M., Foster, T. J. & Prince, A. S. (2006) *Staphylococcus aureus* Protein A Activates TNFR1 Signaling through Conserved IgG Binding Domains. *J Biol Chem* **281**(29): 20190-20196.

Gordon, D. (2004) Viewing and Editing Assembled Sequences Using Consed. In: *Current Protocols in Bioinformatics.* A. D. Baxevanis & D. B. Davison (eds). New York: John Wiley & Co.

Gorwitz, Rachel J., Kruszon-Moran, D., McAllister, Sigrid K., McQuillan, G., McDougal, Linda K., Fosheim, Gregory E., Jensen, Bette J., Killgore, G., Tenover, Fred C. & Kuehnert, Matthew J. (2008) Changes in the Prevalence of Nasal Colonization with *Staphylococcus aureus* in the United States, 2001-2004. *J Infect Dis* **197**(9): 1226-1234.

Goto, H., Shimada, K., Ikemoto, H., Oguri, T. & Study Group on Antimicrobial Susceptibility of Pathogens Isolated from Respiratory Infections (2009) Antimicrobial susceptibility of pathogens isolated from more than 10 000 patients with infectious respiratory diseases: a 25-year longitudinal study. *J Infect Chemother*. **15**(6): 347-360.

Gotz, F. (2002) Microreview on Staphylococcus and biofilms. *Mol Microbiol* **43**: 1367-1378.

Greene, C., McDevitt, D., Francois, P., Vaudaux, P., Lew, D. & Foster, T. (1995) Adhesion properties of mutants of *Staphylococcus aureus* defective in fibronectin-binding proteins and studies on the expression of fnb genes. *Mol Microbiol* **17**(6): 1143-1152.

Guinane, C., Ben Zakour, N., Tormo-Mas, M., Weinert, L., Lowder, B., Cartwright, R., Smyth, D., Smyth, C., Lindsay, J., Gould, K., Witney, A., Hinds, J., Bollback, J., Rambaut, A., Penadés, J. & Fitzgerald, J. (2010) Evolutionary Genomics of *Staphylococcus aureus* Reveals Insights into the Origin and Molecular Basis of Ruminant Host Adaptation. *Genome Biol Evol* **12**(2): 454-466.

Guinane, Caitriona M., Sturdevant, Daniel E., Herron-Olson, L., Otto, M., Smyth, Davida S., Villaruz, Amer E., Kapur, V., Hartigan, Patrick J., Smyth, Cyril J. & Fitzgerald, J. R. (2008) Pathogenomic Analysis of the Common Bovine *Staphylococcus aureus* Clone (ET3): Emergence of a Virulent Subtype with Potential Risk to Public Health. *J Infect Dis* **197**(2): 205-213.

Hageman, J., Pegues, D. & Jepson, C. (2001) Vancomycin-intermediate Staphylococcus aureus in a home health-care patient. . *Emerg Infect Dis* **7**: 1023-1025.

Harris, S. R., Feil, E. J., Holden, M. T. G., Quail, M. A., Nickerson, E. K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J. A., Edgeworth, J. D., de Lencastre, H., Parkhill, J., Peacock, S. J. & Bentley, S. D. (2010) Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* **327**(5964): 469-474.

Hartl, D. L. & Clark, A. G. (2007) *Principles of Population Genetics, Fourth Ed.* Sunderland, MA, USA: Sinauer Associates.

Hartleib, J., Kohler, N., Dickinson, R. B., Chhatwal, G. S., Sixma, J. J., Hartford, O. M., Foster, T. J., Peters, G., Kehrel, B. E. & Herrmann, M. (2000) Protein A is the von Willebrand factor binding protein on *Staphylococcus aureus*. *Blood* **96**(6): 2149-2156.

Hasman, H., Moodley, A., Guardabassi, L., Stegger, M., Skov, R. L. & Aarestrup, F. M. (2010) spa type distribution in *Staphylococcus aureus* originating from pigs, cattle and poultry. *Vet Microbiol* **141**(3-4): 326-331.

Heilmann, C., Niemann, S., Sinha, B., Herrmann, M., Kehrel, B. E. & Peters, G. (2004) *Staphylococcus aureus* fibronectin-binding protein (Fnbp)-mediated adherence to platelets, and aggregation of platelets induced by FnbpA but not by FnbpB. *J Infect Dis* **190**: 321-329.

Hentzer, M., Wu, H., Andersen, J. B., Riedel, K., Rasmussen, T. B., Bagge, N., Kumar, N., Schembri, M. A., Song, Z., Kristoffersen, P., Manefield, M., Costerton, J. W., Molin, S., Eberl, L., Steinberg, P., Kjelleberg, S., Hoiby, N. & Givskov, M. (2003) Attenuation of *Pseudomonas aeruginosa* virulence by quorum sensing inhibitors. *EMBO J* **22**(15): 3803-3815.

Hermans, K., Devriese, L. A. & Haesebrouck, F. (2003) Rabbit staphylococcosis: difficult solutions for serious problems. *Vet Microbiol* **91**(1): 57-64.

Herron-Olson, L., Fitzgerald, J. R., Musser, J. M. & Kapur, V. (2007) Molecular Correlates of Host Specialization in *Staphylococcus aureus*. *PLoS ONE* **2**(10): e1120.

Highlander, S., Hulten, K., Qin, X., Jiang, H., Yerrapragada, S., Mason, E., Shang, Y., Williams, T., Fortunov, R., Liu, Y., Igboeli, O., Petrosino, J., Tirumalai, M., Uzman, A., Fox, G., Cardenas, A., Muzny, D., Hemphill, L., Ding, Y., Dugan, S., Blyth, P., Buhay, C., Dinh, H., Hawes, A., Holder, M., Kovar, C., Lee, S., Liu, W., Nazareth, L., Wang, Q., Zhou, J., Kaplan, S. & Weinstock, G. (2007) Subtle genetic changes enhance virulence of methicillin-resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol* **7**(1): 99.

Hiramatsu, K. (1997) Reduced susceptibility of *Staphylococcus aureus* to vancomycin - Japan, 1996. *MMWR Morb Mortal Wkly Rep* **46**: 624-626.

Hiramatsu, K., Asada, K., Suzuki, E., Okonogi, K. & Yokota, T. (1992) Molecular cloning and nucleotide sequence determination of the regulator region of mecA gene in methicillin-resistant *Staphylococcus aureus* (MRSA). *FEBS Lett* **298**(2-3): 133-136.

Holden, M. T. & Lindsay, J. A. (2008) Whole Genomes: Sequence, Microarray and Systems Biology. In: *Staphylococcus: Molecular Genetics.* J. A. Lindsay (ed). Norfolk, UK: Caister Academic Press.

Holden, M. T. G., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P. J., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G. & Parkhill, J. (2004) Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance. *P Nat Acad Sci USA* **101**(26): 9786-9791.

Holden, M. T. G., Lindsay, J. A., Corton, C., Quail, M. A., Cockfield, J. D., Pathak, S., Batra, R., Parkhill, J., Bentley, S. D. & Edgeworth, J. D. (2010) Genome Sequence of a Recently Emerged, Highly Transmissible, Multi-Antibiotic- and

Antiseptic-Resistant Variant of Methicillin-Resistant *Staphylococcus aureus*, Sequence Type 239 (TW). *J Bacteriol* **192**(3): 888-892.

Hooper, D. C. (2002) Fluoroquinolone resistance among Gram-positive cocci. *The Lancet Infect Dis* **2**(9): 530-538.

Horsburgh, M. J. (2008) The Response of *S. aureus* to Environmental Stimuli. In: *Staphylococcus molecular genetics.* J. A. Lindsay (ed). Norfolk, UK: Caister Academic Press.

Howden, B. P., Davies, J. K., Johnson, P. D. R., Stinear, T. P. & Grayson, M. L. (2010) Reduced Vancomycin Susceptibility in *Staphylococcus aureus*, Including Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate Strains: Resistance Mechanisms, Laboratory Detection, and Clinical Implications. *Clin Microbiol Rev* **23**(1): 99-139.

Howden, B. P., Stinear, T. P., Allen, D. L., Johnson, P. D. R., Ward, P. B. & Davies, J. K. (2008) Genomic Analysis Reveals a Point Mutation in the Two-Component Sensor Gene graS That Leads to Intermediate Vancomycin Resistance in clinical *Staphylococcus aureus*. *Antimicrob Agents Ch* **52**(10): 3755-3762.

Humbert, O., Prudhomme, M., Hakenbeck, R., Dowson, C. G. & Claverys, J. P. (1995) Homologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *P Natl Acad Sci USA.* **92**(20): 9052–9056.

Iandolo, J. J., Worrell, V., Groicher, K. H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., Qi, S., Zhu, H. & Roe, B. A. (2002) Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene* **289**(1-2): 109-118.

Isigidi, B. K., Devriese, L. A., Godard, C. & Vanhoof, J. (1990) Characteristics of *Staphylococcus aureus* Associated with Meat-Products and Meat Workers. *Lett Appl Microbiol* **11**(3): 145-147.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C. & Hiramatsu, K. (2001) Structural comparison of three types of staphylococcal

cassette chromosome mec integrated into the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Ch* **45**: 1323-1336.

Ito, T., Katayama, Y. & Hiramatsu, K. (1999) Cloning and Nucleotide Sequence Determination of the Entire mec DNA of Pre-Methicillin-Resistant *Staphylococcus aureus* N315. *Antimicrob Agents Ch* **43**(6): 1449-1458.

Jones, R. N. (2003) Global epidemiology of antimicrobial resistance among community-acquired and nosocomial pathogens: a five-year summary from the SENTRY Antimicrobial Surveillance Program (1997-2001). *Semin Respir Crit Care Med.* **24**(1): 121-134.

Kang, W. & Reid, K. B. M. (2003) DMBT1, a regulator of mucosal homeostasis through the linking of mucosal defense and regeneration? *FEBS Letters* **540**(1-3): 21-25.

Kappes, R. M., Kempf, B. & Bremer, E. (1996) Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD. *J Bacteriol* **178**(17): 5071-5079.

Kapral, F. A., Smith, S. & Lal, D. (1992) The esterification of fatty acids by *Staphylococcus aureus* fatty acid modifying enzyme (FAME) and its inhibition by glycerides. *J Med Microbiol* **37**(4): 235-237.

Katayama, Y., Ito, T. & Hiramatsu, K. (2000) A New Class of Genetic Element, Staphylococcus Cassette Chromosome mec, Encodes Methicillin Resistance in *Staphylococcus aureus*. *Antimicrob Agents Ch* **44**(6): 1549-1555.

Kew, O. M., Nottay, B. K., Hatch, M. H., Hierholzer, J. C. & Obijeski, J. F. (1983) Oligonucleotide fingerprint analysis of enterovirus 70 isolates from the 1980 to 1981 pandemic of acute hemorrhagic conjunctivitis: evidence for a close genetic relationship among Asian and American strains. *Infect Immun* **41**(2): 631-635.

Khan, S. & Novick, R. (1983) Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from *Staphylococcus aureus*. *Plasmid* **10**(3): 251-259.

Khan, S. A. (2005) Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* **53**: 126-136

Khurshid, M., Chou, T., Carey, R., Larsen, R., Conover, C. & Bornstein, S. (2000) *Staphylococcus aureus* with reduced susceptibility to vancomycin - Illinois, 1999. *MMWR Morb Mortal Wkly Rep* **48**: 1165-1167.

King, M. D., Humphrey, B. J., Wang, Y. F., Kourbatova, E. V., Ray, S. M. & Blumberg, H. M. (2006) Emergence of Community-Acquired Methicillin-Resistant *Staphylococcus aureus* USA 300 Clone as the Predominant Cause of Skin and Soft-Tissue Infections. *Ann Intern Med* **144**(5): 309-317.

Knobler, S., Mahmoud, A., Lemon, S., Mack, A., Sivitz, L. & Oberholtzer, K. (2004) *LEARNING FROM SARS: Preparing for the Next Disease Outbreak.* Washington, D.C.: The National Academies Press.

Kreiswirth, B., Kornblum, J., Arbeit, R., Eisner, W., Maslow, J., McGeer, A., Low, D. & Novick, R. (1993) Evidence for a clonal origin of methicillin resistance in *Staphylococcus aureus*. *Science* **259**(5092): 227-230.

Kullik, I. & Giachino, P. (1997) The alternative sigma factor σB in *Staphylococcus aureus*: regulation of the sigB operon in response to growth phase and heat shock *Arch Microbiol* **167**(2/3): 151-159.

Kuramasu, S., Imamura, Y., Takizawa, T., Oguchi, F. & Tajima, Y. (1967) Studies on staphylococcosis in young chickens I. Outbreaks of staphylococcal infection on poultry farms and characteristics of *Staphylococcus aureus* isolated from chickens. *Zbl Vet Med* **B14**: 646-656.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K.-i., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N. K., Sawano, T., Inoue, R.-i., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. & Hiramatsu, K. (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *The Lancet* **357**(9264): 1225-1240.

Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. (2005) The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *P Nat Acad Sci USA* **102**(14): 5174-5179.

Ladhani, S. (2003) Understanding the mechanism of action of the exfoliative toxins of *Staphylococcus aureus*. *FEMS Immunol Med Micro* **39**(2): 181-189.

Lannergard, J., Norstrom, T. & Hughes, D. (2009) Genetic Determinants of Resistance to Fusidic Acid among Clinical Bacteremia Isolates of *Staphylococcus aureus*. *Antimicrob Agents Ch* **53**(5): 2059-2065.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007) ClustalW and ClustalX version 2. *Bioinformatics* **23**(21): 2947-2948.

Lawrence, J. G. (2006) Studying Evolution Using Genome Sequence Data. In: *Evolution of Microbial Pathogens.* S. Seifert & V. J. DiRita (eds). Washington, DC, USA: ASM Press.

Lebarbenchon, C., Brown, S. P., Poulin, R., Gauthier-Clerc, M. & Thomas, F. (2008) Evolution of pathogens in a man-made world. *Mol Ecol* **17**(1): 475-484.

Lee, J. S. & Stewart, P. R. (1985) The Virion Proteins and Ultrastructure of *Staphylococcus aureus* Bacteriophages. *J Gen Virol* **66**(9): 2017-2027.

Lee, P.K., Kreiswirth, B.N., Deringer, J.R., Projan, S.J., Eisner, W., Smith, B.L., Carlson, E., Novick, R.P., Schlievert, P.M. (1992) Nucleotide sequences and biologic properties of toxic shock syndrome toxin 1 from ovine- and bovine-associated *Staphylococcus aureus*. *J Infect Dis* **165**: 1056–1063.

Lefebure, T. & Stanhope, M. J. (2007) Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. *Genome Biology* **8**(5): R71.

Lefebure, T. & Stanhope, M. J. (2009) Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus Campylobacter. *Genome Res.* **19**: 1224-1232

Leonard, F. C. & Markey, B. K. (2008) Meticillin-resistant *Staphylococcus aureus* in animals: A review. *Vet J* **175**(1): 27-36.

Lerat, E. & Ochman, H. (2005) Recognizing the pseudogenes in bacterial genomes. *Nucl Acids Res* **33**(10): 3125-3132.

Leski, T., Oliveira, D., Trzcinski, K., Sanches, I. S., de Sousa, M. A., Hryniewicz, W. & de Lencastre, H. (1998) Clonal Distribution of Methicillin-Resistant *Staphylococcus aureus* in Poland. *J Clin Microbiol* **36**(12): 3532-3539.

Lindsay, J. A. (2008) *S. aureus* Evolution: Lineages and Mobile Genetic Elements (MGEs). In: *Staphylococcus: Molecular Genetics.* J. A. Lindsay (ed). Norfolk, UK: Caister Academic Press.

Lindsay, J. A. & Holden, M. T. G. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct Integr Genomics* **6**(3): 186-201.

Lindsay, J. A., Moore, C. E., Day, N. P., Peacock, S. J., Witney, A. A., Stabler, R. A., Husain, S. E., Butcher, P. D. & Hinds, J. (2006) Microarrays Reveal that Each of the Ten Dominant Lineages of *Staphylococcus aureus* Has a Unique Combination of Surface-Associated and Regulatory Genes. *J Bacteriol* **188**(2): 669-676.

Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus. Molecular Microbiology* **29**(2): 527-543.

Liu, Y., Harrison, P., Kunin, V. & Gerstein, M. (2004) Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol* **5**(9): R64.

Lowder, B. V., Guinane, C. M., Ben Zakour, N. L., Weinert, L. A., Conway-Morris, A., Cartwright, R. A., Simpson, A. J., Rambaut, A., Nubel, U. & Fitzgerald, J. R. (2009) Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *P Nat Acad Sci USA* **106**(46): 19545-19550.

Lucht, J. & Bremer, E. (1994) Adaptation of *Escherichia coli* to high osmolarity environments: osmoregulation of the high-affinity glycine betaine transport system proU. *FEMS Microbiol Rev* **14**(1): 3-20.

Luong, T. T., Ouyang, S., Bush, K. & Lee, C. Y. (2002) Type 1 Capsule Genes of *Staphylococcus aureus* Are Carried in a Staphylococcal Cassette Chromosome Genetic Element. *J Bacteriol* **184**(13): 3623-3629.

Ministry of Agriculture, Fisheries and Food (1981) Leg Weakness in Poultry: The Report of a Ministry of Agriculture, Fisheries and Food (MAFF) working party.

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *P Nat Acad Sci USA* **95**(6): 3140-3145.

Maiden, M. C. J. & Urwin, R. (2006) Population Dynamics of Bacterial Pathogens. In: *Evolution of Microbial Pathogens.* S. Seifert & V. J. DiRita (eds). Washington, DC, USA: ASM Press.

Maione, D., Margarit, I., Rinaudo, C. D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E. T., Rosini, R., D'Agostino, N., Miorin, L., Buccato, S., Mariani, M., Galli, G., Nogarotto, R., Dei, V. N., Vegni, F., Fraser, C., Mancuso, G., Teti, G., Madoff, L. C., Paoletti, L. C., Rappuoli, R., Kasper, D. L., Telford, J. L. & Grandi, G. (2005) Identification of a Universal Group B Streptococcus Vaccine by Multiple Genome Screen. *Science* **309**(5731): 148-150.

Makris, G., Wright, J. D., Ingham, E. & Holland, K. T. (2004) The hyaluronate lyase of *Staphylococcus aureus* - a virulence factor? *Microbiology* **150**(6): 2005-2013.

Mandel, M. J., Wollenberg, M. S., Stabb, E. V., Visick, K. L. & Ruby, E. G. (2009) A single regulatory gene is sufficient to alter bacterial host range. *Nature* **458**(7235): 215-218.

Manna, A. C. & Cheung, A. L. (2006) Expression of SarX, a Negative Regulator of agr and Exoprotein Synthesis, Is Activated by MgrA in *Staphylococcus aureus*. *J Bacteriol* **188**(12): 4288-4299.

Manning, L. & Baines, R. N. (2004) Globalisation: a study of the poultry-meat supply chain. *Br Food J* **106**(10/11): 819-836.

Massey, R. C., Horsburgh, M. J., Lina, G., Hook, M. & Recker, M. (2006) The evolution and maintenance of virulence in *Staphylococcus aureus*: a role for host-to-host transmission? *Nat Rev Microbiol* **4**(12): 953-958.

Massimi, I., Park, E., Rice, K., Muller-Esterl, W., Sauder, D. & McGavin, M. J. (2002) Identification of a Novel Maturation Mechanism and Restricted Substrate Specificity for the SspB Cysteine Protease of *Staphylococcus aureus*. *J Biol Chem* **277**(44): 41770-41777.

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J. M., Koehrsen, M., Rokas, A., Yandava, C. N., Engels, R., Zeng, E., Olavarietta, R., Doud, M., Smith, R. S., Montgomery, P., White, J. R., Godfrey, P. A., Kodira, C., Birren, B., Galagan, J. E. & Lory, S. (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *P Nat Acad Sci USA* **105**(8): 3100-3105.

Matthews, J. (1999) *Diseases of the Goat*. Oxford, UK: Blackwell Science.

McDevitt, D., Nanavatay, T., House-Pompeo, K., Bell, E., Turner, N., McIntire, L., Foster, T. & Hook, M. (1997) Characterization of the interaction between the *Staphylococcus aureus* clumping factor A (ClfA) and fibrinogen. *Eur J Biochem* **247**: 416-424.

McDevitt, D., Vaudaux, P. & Foster, T. J. (1992) Genetic evidence that bound coagulase of *Staphylococcus aureus* is not clumping factor. *Infect Immun* **60**(4): 1514-1523.

McNamee, P. T., McCullagh, J. J., Thorp, B. H., Ball, H. J., Graham, D., McCullough, S. J., McConaghy, D. & Smyth, J. A. (1998) Study of leg weakness in two commercial broiler flocks. *Vet Rec* **143**(5): 131-135.

McNamee, P. T. & Smyth, J. A. (2000) Bacterial chondronecrosis with osteolmyelitis ('femoral head necrosis') of broiler chickens: a review. *Avian Pathol* **29**: 253-270.

McNeil, P. E. (2007) Staphylococcal skin infections. In: *Diseases of Sheep.* A. D. Aitken (ed). Oxford, UK: Blackwell Publishing.

McNeill, W. (1989) *Plagues and Peoples, 2nd Edition*. New York, USA: Anchor Books.

McShan, W. M. & Ferretti, J. J. (2007) Prophages and Their Contribution to Host Cell Phenotype. In: *Bacteriophage: Genetics and Molecular Biology.* S. McGrath & D. van Sinderen (eds). Norfolk, UK: Caister Academic Press.

Meats, E., Feil, E. J., Stringer, S., Cody, A. J., Goldstein, R., Kroll, J. S., Popovic, T. & Spratt, B. G. (2003) Characterization of Encapsulated and Noncapsulated *Haemophilus influenzae* and Determination of Phylogenetic Relationships by Multilocus Sequence Typing. *J Clin Microbiol* **41**(4): 1623-1636.

Melles, D. C., Tenover, F. C., Kuehnert, M. J., Witsenboer, H., Peeters, J. K., Verbrugh, H. A. & van Belkum, A. (2008) Overlapping Population Structures of Nasal Isolates of *Staphylococcus aureus* from Healthy Dutch and American Individuals. *J Clin Microbiol* **46**(1): 235-241.

Mempel, M., Schmidt, T., Weidinger, S., Schnopp, C., Foster, T., Ring, J. & Abeck, D. (1998) Role of *Staphylococcus Aureus* Surface-Associated Proteins in the Attachment to Cultured HaCaT Keratinocytes in a New Adhesion Assay. *J Invest Dermatol* **111**(3): 452-456.

Mérino, D., Réglier-Poupet, H., Berche, P. & Charbit, A. (2002) A hypermutator phenotype attenuates the virulence of *Listeria monocytogenes* in a mouse model. *Mol Microbiol* **44**(3): 877-887.

Miller, R. A. & Britigan, B. E. (1997) Role of oxidants in microbial pathophysiology. *Clin Microbiol Rev* **10**(1): 1-18.

Monecke, S., Kuhnert, P., Hotzel, H., Slickers, P. & Ehricht, R. (2007) Microarray based study on virulence-associated genes and resistance determinants of *Staphylococcus aureus* isolates from cattle. *Veterinary Microbiology* **125**(1-2): 128-140.

Montoya, M. & Gouaux, E. (2003) β-Barrel membrane protein folding and structure viewed through the lens of α-hemolysin. *BBA Biomembranes* **1609**(1): 19-27.

Moodley, A., Stegger, M., Bagcigil, A. F., Baptiste, K. E., Loeffler, A., Lloyd, D. H., Williams, N. J., Leonard, N., Abbott, Y., Skov, R. & Guardabassi, L. (2006) spa typing of methicillin-resistant *Staphylococcus aureus* isolated from domestic animals and veterinary staff in the UK and Ireland. *J Antimicrob Ch* **58**(6): 1118-1123.

Moran, N. A., McLaughlin, H. J. & Sorek, R. (2009) The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. *Science* **323**(5912): 379-382.

Mork, T., Tollersrud, T., Kvitle, B., Jorgensen, H. J. & Waage, S. (2005) Comparison of *Staphylococcus aureus* Genotypes Recovered from Cases of Bovine, Ovine, and Caprine Mastitis. *J Clin Microbiol* **43**(8): 3979-3984.

Mortensen, J. E., Shryock, T. R. & Kapral, F. A. (1992) Modification of bactericidal fatty acids by an enzyme of *Staphylococcus aureus*. *J Med Microbiol* **36**(4): 293-298.

Mourad, W., Scholl, P., Diaz, A., Geha, R. & Chatila, T. (1989) The staphylococcal toxic shock syndrome toxin 1 triggers B cell proliferation and differentiation via major histocompatibility complex- unrestricted cognate T/B cell interaction. *J Exp Med* **170**: 2011-2022.

Musser, J. M. & Kapur, V. (1992) Clonal analysis of methicillin-resistant *Staphylococcus aureus* strains from intercontinental sources: association of the mec gene with divergent phylogenetic lineages implies dissemination by horizontal transfer and recombination. *J. Clin. Microbiol.* **30**(8): 2058-2063.

Musser, J. M., Schlievert, P. M., Chow, A. W., Ewan, P., Kreiswirth, B. N., Rosdahl, V. T., Naidu, A. S., Witte, W. & Selander, R. K. (1990) A single clone of *Staphylococcus aureus* causes the majority of cases of toxic shock syndrome. *P Nat Acad Sci USA* **87**(1): 225-229.

Mutalib, A., Riddell, C. & Osborne, A. D. (1983) Studies on the Pathogenesis of Staphylococcal Osteomyelitis in Chickens. *Avian Diseases* **27**(1): 141-156.

Mwangi, M. M., Wu, S. W., Zhou, Y., Sieradzki, K., de Lencastre, H., Richardson, P., Bruce, D., Rubin, E., Myers, E., Siggia, E. D. & Tomasz, A. (2007) Tracking the

in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *P Nat Acad Sci USA* **104**(22): 9451-9456.

Nairn, M. E. & Watson, A. R. A. (1972) Leg weakness in poultry - a clinical and pathological characterisation. *Aust Vet Journal* **48**: 645-656.

Narita, S., Kaneko, J., Chiba, J.-i., Piémont, Y., Jarraud, S., Etienne, J. & Kamio, Y. (2001) Phage conversion of Panton-Valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a PVL-converting phage, phi SLT. *Gene* **268**(1-2): 195-206.

Nathanson, N., Wilesmith, J. & Griot, C. (1997) Bovine Spongiform Encephalopathy (BSE): Causes and Consequences of a Common Source Epidemic. *Am J Epidemiol* **145**(11): 959-969.

Neoh, H.-m., Cui, L., Yuzawa, H., Takeuchi, F., Matsuo, M. & Hiramatsu, K. (2008) Mutated Response Regulator graR Is Responsible for Phenotypic Conversion of *Staphylococcus aureus* from Heterogeneous Vancomycin-Intermediate Resistance to Vancomycin-Intermediate Resistance. *Antimicrob Agents Ch* **52**(1): 45-53.

Ní Eidhin, D., Perkins, S., Francois, P., Vaudaux, P., Hook, M. & Foster, T. J. (1998) Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. *Mol Microbiol* **30**: 245-257.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J., Sninsky, J., Adams, M. D. & Cargill, M. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**: e170.

Nielsen, R. & Yang, Z. (1998) Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics* **148**(3): 929-936.

Novick, R. (2006) Staphylococcal Pathogenesis and Pathogenicity Factors: Genetics and Regulation. In: *Gram-Positive Pathogens.* V. A. Fischetti, R. Novick, J. Ferretti, D. Portnoy & J. Rood (eds). Washington, DC: ASM Press.

Nubel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., Žemličková, H., Leblois, R., Wirth, T., Jombart, T., Balloux, F. & Witte, W. (2010) A Timescale for Evolution, Population Expansion, and Spatial Spread of an Emerging Clone of Methicillin-Resistant *Staphylococcus aureus*. *PLoS Pathog* **6**(4).

Nubel, U., Roumagnac, P., Feldkamp, M., Song, J.-H., Ko, K. S., Huang, Y.-C., Coombs, G., Ip, M., Westh, H., Skov, R., Struelens, M. J., Goering, R. V., Strommenger, B., Weller, A., Witte, W. & Achtman, M. (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *P Nat Acad Sci USA* **105**(37): 14130-14135.

O'Riordan, K. & Lee, J. C. (2004) *Staphylococcus aureus* Capsular Polysaccharides. *Clin Microbiol Rev* **17**(1): 218-234.

O'Seaghdha, M., van Schooten, C. J., Kerrigan, S. W., Emsley, J., Silverman, G. J., Cox, D., Lenting, P. J. & Foster, T. J. (2006) *Staphylococcus aureus* protein A binding to von Willebrand factor A1 domain is mediated by conserved IgG binding regions. *Febs J* **273**(21): 4831-4841.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**(6784): 299-304.

Ohta, T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* **246**: 96-98.

Okuma, K., Iwakawa, K., Turnidge, J. D., Grubb, W. B., Bell, J. M., O'Brien, F. G., Coombs, G. W., Pearman, J. W., Tenover, F. C., Kapi, M., Tiensasitorn, C., Ito, T. & Hiramatsu, K. (2002) Dissemination of New Methicillin-Resistant *Staphylococcus aureus* Clones in the Community. *J Clin Microbiol* **40**(11): 4289-4294.

Orsi, R., Sun, Q. & Wiedmann, M. (2008) Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. *BMC Evol Biol* **8**: 233.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R.,

Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O. & Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**(17): 5691-5702.

Pantucek , R., Doskar, J., Kasparek, P., Oracova, E., Kvardova, V. & Rosypal, S. (2004) Identification of bacteriophage types and their carriage in *Staphylococcus aureus*. *Arch Virol* **149**: 1689-1703.

Pearce-Duvet, J. M. C. (2006) The origin of human pathogens: evaluating the role of agriculture and domestic animals in the evolution of human disease. *Biol Rev* **81**(3): 369-382.

Persoons, D., Hoorebeke, S. V., Hermans, K., Butaye, P., de Kruif, A., Haesebrouck, F. & Dewulf, J. (2009) Methicillin-Resistant *Staphylococcus aureus* in Poultry. *Emerging Infectious Diseases* **15**(3): 452-453.

Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M. & Nielsen, R. (2007) Genes under positive selection in *Escherichia coli*. *Genome Res* **17**(9): 1336-1343.

Pond, S. L., Frost, S. D. & Muse, S. V. (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**: 676-679.

Postma, P. W., Lengeler, J. W. & Jacobson, G. R. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57**(3): 543-594.

Projan, S. J. & Novick, R. P. (1997) The Molecular Basis of Pathogenicity. In: *The Staphylococci in Human Disease.* K. B. Crossley & G. L. Archer (eds). Edinburgh, UK: Churchill Livingstone Inc.

Prudhomme, M., Méjean, V., Martin, B. & Claverys, J. P. (1991) Mismatch repair genes of *Streptococcus pneumoniae*: HexA confers a mutator phenotype in *Escherichia coli* by negative complementation. *J Bacteriol* **173**(22): 7196–7203.

Quinn, P. J., Markey, B. K., Carter, M. E., Donnelly, W. J. & Leonard, F. C. (2002) *Veterinary Microbiology and Microbial Disease*. Blackwell Science Ltd., London.

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K. & Holmes, E. C. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**(7195): 615-619.

Ratledge, C. & Dover, L. G. (2000) Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* **54**(1): 881-941.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., Couronne, O., Hua, S., Smith, M. A., Zhang, P., Liu, J., Bussemaker, H. J., van Batenburg, M. F., Howells, S. L., Scherer, S. E., Sodergren, E., Matthews, B. B., Crosby, M. A., Schroeder, A. J., Ortiz-Barrientos, D., Rives, C. M., Metzker, M. L., Muzny, D. M., Scott, G., Steffen, D., Wheeler, D. A., Worley, K. C., Havlak, P., Durbin, K. J., Egan, A., Gill, R., Hume, J., Morgan, M. B., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K. P., Dubchak, I., Noor, M. A., Anderson, W., White, K. P., Clark, A. G., Schaeffer, S. W., Gelbart, W., Weinstock, G. M. & Gibbs, R. A. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Res* **15**: 1-18.

Rocha, E. P., Smith, J. M., Hurst, L. D., Holden, M. T., Cooper, J. E., Smith, N. H., Feil, E. J. (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**: 226-235.

Roche, F. M., Downer, R., Keane, F., Speziale, P., Park, P. W. & Foster, T. J. (2004) The N-terminal A domain of fibronectin binding proteins A and B promotes adhesion of *Staphylococcus aureus* to elastin. *J Biol Chem* **279**: 38433-38440.

Roche, F. M., Massey, R., Peacock, S. J., Day, N. P. J., Visai, L., Speziale, P., Lam, A., Pallen, M. & Foster, T. J. (2003) Characterization of novel LPXTG-containing proteins of *Staphylococcus aureus* identified from genome sequences. *Microbiol* **149**(3): 643-654.

Rodgers, J. D., McCullagh, J. J., McNamee, P. T., Smyth, J. A. & Ball, H. J. (1999) Comparison of *Staphylococcus aureus* recovered from personnel in a poultry hatchery and in broiler parent farms with those isolated from skeletal disease in broilers. *Vet Microbiol* **69**: 189-198.

Rolain, J.-M., Francois, P., Hernandez, D., Bittar, F., Richet, H., Fournous, G., Mattenberger, Y., Bosdure, E., Stremler, N., Dubus, J.-C., Sarles, J., Reynaud-Gaubert, M., Boniface, S., Schrenzel, J. & Raoult, D. (2009) Genomic analysis of an emerging multiresistant *Staphylococcus aureus* strain rapidly spreading in cystic fibrosis patients revealed the presence of an antibiotic inducible bacteriophage. *Biol Direct* **4**(1): 1.

Rotun, S. S., McMath, V., Schoonmaker, D. J., Maupin, P. S., Tenover, F. C., Hill, B. C. & Ackman, D. M. (1999) Staphylococcus aureus with reduced susceptibility to vancomycin isolated from a patient with fatal bacteremia. *Emerg Infect Dis* **5**(1): 147-149.

Ruimy, R., Armand-Lefevre, L., Barbier, F., Ruppe, E., Cocojaru, R., Mesli, Y., Maiga, A., Benkalfat, M., Benchouk, S., Hassaine, H., Dufourcq, J.-B., Nareth, C., Sarthou, J.-L., Andremont, A. & Feil, E. J. (2009) Comparisons between Geographically Diverse Samples of Carried *Staphylococcus aureus*. *J Bacteriol* **191**(18): 5577-5583.

Ruimy, R., Maiga, A., Armand-Lefevre, L., Maiga, I., Diallo, A., Koumare, A. K., Ouattara, K., Soumare, S., Gaillard, K., Lucet, J.-C., Andremont, A. & Feil, E. J. (2008) The Carriage Population of *Staphylococcus aureus* from Mali Is Composed of a Combination of Pandemic Clones and the Divergent Panton-Valentine Leukocidin-Positive Genotype ST152. *J. Bacteriol.* **190**(11): 3962-3968.

Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Shaw, M. W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R. A. M. & Smith, D. J.

(2008) The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science* **320**(5874): 340-346.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Ruzin, A., Lindsay, J. & Novick, R. P. (2001) Molecular genetics of SaPI1 - a mobile pathogenicity island in *Staphylococcus aureus*. *Molecular Microbiology* **41**(2): 365-377.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.

Sancak, B., Ercis, S., Menemenlioglu, D., Colakoglu, S. & Hascelik, G. (2005) Methicillin-resistant *Staphylococcus aureus* heterogeneously resistant to vancomycin in a Turkish university hospital. *J Antimicrob Ch* **56**: 519-523.

Schmidt, K. A., Manna, A. C., Gill, S. & Cheung, A. L. (2001) SarT, a repressor of alpha-hemolysin in *Staphylococcus aureus*. *Infect Immun* **69**(8): 4749-4758.

Schmiel, D. H. & Miller, V. L. (1999) Bacterial phospholipases and pathogenesis. *Microbes Infect* **1**(13): 1103-1112.

Schneewind, O., Mihaylova-Petkov, D. & Model, P. (1993) Cell wall sorting signals in surface proteins of Gram-positive bacteria. *EMBO J* **12**: 4803-4811.

Senn, M. M., Giachino, P., Homerova, D., Steinhuber, A., Strassner, J., Kormanec, J., Fluckiger, U., Berger-Bachi, B. & Bischoff, M. (2005) Molecular Analysis and Organization of the sigma B Operon in *Staphylococcus aureus*. *J Bacteriol* **187**(23): 8006-8019.

Seybold, U., Kourbatova, Ekaterina V., Johnson, James G., Halvosa, Sue J., Wang, Yun F., King, Mark D., Ray, Susan M. & Blumberg, Henry M. (2006) Emergence of Community Associated Methicillin-Resistant *Staphylococcus aureus* USA300 Genotype as a Major Cause of Health Care Associated Blood Stream Infections. *Clin Infect Dis* **42**(5): 647-656.

Shands, K., Schmid, G., Dan, B., Blum, D., Guidotti, R., Hargrett, N., Anderson, R., Hill, D., Broome, C., Band, J. & Fraser, D. (1980) Toxic-shock syndrome in

menstruating women: association with tampon use and *Staphylococcus aureus* and clinical features in 52 cases. *N Engl J Med* **303**(25): 1436-1442.

Shen, Y., Liu, J., Estiu, G., Isin, B., Ahn, Y. Y., Lee, D. S., Barabasi, A. L., Kapatral, V., Wiest, O. & Oltvai, Z. N. (2010) Blueprint for antimicrobial hit discovery targeting metabolic networks. *P Nat Acad Sci USA* **107**(3): 1082-1087.

Shriner, D., Nickle, D. C., Jensen, M. A. & Mullins, J. I. (2003) Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genet Res* **81**: 115-121.

Siboo, I. R., Chaffin, D. O., Rubens, C. E. & Sullam, P. M. (2008) Characterization of the Accessory Sec System of *Staphylococcus aureus*. *J Bacteriol* **190**(18): 6188-6196.

Sischo, W. M., Heider, L. E., Miller, G. Y. & Moore, D. A. (1993) Prevalence of contagious pathogens of bovine mastitis and use of mastitis control practices. *J Am Vet Med Assoc* **202**(4): 595-600.

Skaar, E. & Schneewind, O. (2004) Iron-regulated surface determinants (Isd) of *Staphylococcus aureus*: stealing iron from heme. *Microb Infect* **6**: 390-397.

Smith, E. M., Green, L. E., Medley, G. F., Bird, H. E., Fox, L. K., Schukken, Y. H., Kruze, J. V., Bradley, A. J., Zadoks, R. N. & Dowson, C. G. (2005) Multilocus Sequence Typing of Intercontinental Bovine *Staphylococcus aureus* Isolates. *J. Clin. Microbiol.* **43**(9): 4737-4743.

Smith, H. W. (1954) Experimental Staphylococcal Infection in Chickens. *J Path Bact* **67**: 81-87.

Smith, P. A. & Romesberg, F. E. (2007) Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nat Chem Biol* **3**(9): 549-556.

Smyth, D. S., Feil, E. J., Meaney, W. J., Hartigan, P. J., Tollersrud, T., Fitzgerald, J. R., Enright, M. C. & Smyth, C. J. (2009) Molecular genetic typing reveals further insights into the diversity of animal-associated *Staphylococcus aureus*. *J Med Microbiol* **58**(10): 1343-1353.

Smyth, J. A. & McNamee, P. T. (2008) Staphylococci, Streptococci and Enterococci. In: *Poultry diseases, 6th edition.* F. Jordan, M. Pattison, D. Alexander & T. Faragher (eds). Elsevier Health Sciences.

Soyer, Y., Orsi, R., Rodriguez-Rivera, L., Sun, Q. & Wiedmann, M. (2009) Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected Salmonella serotypes. *BMC Evolutionary Biology* **9**(1): 264.

Staden, R., Judge, D. & Bonfield, J. (1998) Sequence assembly and finishing methods. In: *Bioinformatics. A Practical Guide to the Analysis of Genes and Proteins, 2nd ed.* A. Baxevanis & B. Ouellette (eds). New York, USA: John Wiley & Sons.

Staeheli, P., Puehler, F., Schneider, K., Göbel, T. W. & Kaspers, B. (2001) Cytokines of Birds: Conserved Functions - A Largely Different Look. *J Interf Cytok Res* **21**(12): 993-1010.

Stapleton, P. D. & Taylor, P. W. (2002) Methicillin resistance in *Staphylcoccus aureus*: mechanisms and modulation. *Sci Prog* **85**(1): 5-72.

Stearns, S. C. & Hoekstra, R. F. (2002) *Evolution: an introduction*. New York, USA: Oxford University Press Inc.

Stewart, P. R., Waldron, H. G., Lee, J. S. & Matthews, P. R. (1985) Molecular relationships among serogroup B bacteriophages of *Staphylococcus aureus. J Virol.* **55**(1): 111-116.

Stranger-Jones, Y. K., Taeok, B. & Schneewind, O. (2006) Vaccine assembly from surface proteins of *Staphylococcus aureus*. *P Nat Acad Sci USA* **103**(45): 16942–16947.

Strommenger, B., Kehrenberg, C., Kettlitz, C., Cuny, C., Verspohl, J., Witte, W. & Schwarz, S. (2006) Molecular characterization of methicillin-resistant *Staphylococcus aureus* strains from pet animals and their relationship to human isolates. *J Antimicrob Ch* **57**(3): 461-465.

Subedi, A., Ubeda, C., Adhikari, R. P., Penades, J. R. & Novick, R. P. (2007) Sequence analysis reveals genetic exchanges and intraspecific spread of SaPI2, a pathogenicity island involved in menstrual toxic shock. *Microbiology* **153**(10): 3235-3245.

Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I. & Achtman, M. (1998) Free recombination within *Helicobacter pylori*. *P Nat Acad Sci USA* **95**(21): 12619-12624.

Sung, J. M. L. & Lindsay, J. A. (2007) *Staphylococcus aureus* Strains that are Hypersusceptible to Resistance Gene Transfer from Enterococci. *Antimicrob Agents Ch* **51**(6): 2189-2191.

Sung, J. M. L., Lloyd, D. H. & Lindsay, J. A. (2008) *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* **154**(7): 1949-1959.

Takeuchi, O. & Akira, S. (2010) Pattern Recognition Receptors and Inflammation. *Cell* **140**(6): 805-820.

Takeuchi, S., Kinoshita, T., Kaidoh, T. & Hashizume, N. (1999) Purification and characterization of protease produced by *Staphylococcus aureus* isolated from a diseased chicken. *Vet Microbiol* **67**(3): 195-202.

Takeuchi, S., Matsunaga, K., Inubushi, S., Higuchi, H., Imaizumi, K. & Kaidoh, T. (2002) Structural gene and strain specificity of a novel cysteine protease produced by *Staphylococcus aureus* isolated from a diseased chicken. *Vet Microbiol* **89**: 201-210.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.

Tanaka, M., Wang, T., Onodera, Y., Uchida, Y. & Sato, K. (2000) Mechanism of quinolone resistance in *Staphylococcus aureus*. *J Infect Chemother* **6**: 131-139.

Taylor, D. J. (1999) *Pig Diseases, Seventh Edition*. Suffolk, UK: St Edmundsbury Press.

Tcherepanov, T., Ehlers, A. & Upton, C. (2006) Genome Annotation Transfer Utility (GATU): Rapid annotation of viral genomes using a closely related reference genome. *BMC Genomics* **7**: 150.

Tenover, F., McDougal, L., Goering, R., Killgore, G., Projan, S. J., Patel, J. B. & Dunman, P. M. (2006) Characterisation of a strain of community-associated methicillin-resistant *Staphylococcus aureus* widely disseminated in the United States. *J Clin Microbiol* **44**: 108-118.

Thomson, N. R., Clayton, D. J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., Quail, M. A., Stevens, M., Jones, M. A., Watson, M., Barron, A., Layton, A., Pickard, D., Kingsley, R. A., Bignell, A., Clark, L., Harris, B., Ormond, D., Abdellah, Z., Brooks, K., Cherevach, I., Chillingworth, T., Woodward, J., Norberczak, H., Lord, A., Arrowsmith, C., Jagels, K., Moule, S., Mungall, K., Sanders, M., Whitehead, S., Chabalgoity, J. A., Maskell, D., Humphrey, T., Roberts, M., Barrow, P. A., Dougan, G. & Parkhill, J. (2008) Comparative genome analysis of *Salmonella Enteritidis* PT4 and *Salmonella Gallinarum* 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**(10): 1624-1637.

Todd, J., Fishaut, M., Kapral, F. & Welch, T. (1978) Toxic Shock Syndrome associated with Phage Group I Staphylococci. *The Lancet* **312**(8100): 1116-1118.

Tormo, M. A., Ferrer, M. D., Maiques, E., Ubeda, C., Selva, L., Lasa, I., Calvete, J. J., Novick, R. P. & Penades, J. R. (2008) *Staphylococcus aureus* Pathogenicity Island DNA Is Packaged in Particles Composed of Phage Proteins. *J. Bacteriol.* **190**(7): 2434-2440.

Townsend, D. E., Kaenjak, A., Jayaswal, R. K. & Wilkinson, B. J. (1996) Proline is biosynthesized form arginine in *Staphylococcus aureus*. *Microbiol* **142**(6): 1491-1497.

Trzciński, K., van Leeuwen, W., van Belkum, A., Grzesiowski, P., Kluytmans, J., Sijmons, M., Verbrugh, H., Witte, W. & Hryniewicz, W. (1997) Two clones of methicillin-resistant *Staphylococcus aureus* in Poland. *Clin Microbiol Infect* **3**(2): 198-207.

Úbeda, C., Barry, P., Penades, J. R. & Novick, R. P. (2007a) A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *P Nat Acad Sci USA* **104**(36): 14182-14188.

Úbeda, C., Maiques, E., Knecht, E., Lasa, Í., Novick, R. P. & Penadés, J. R. (2005) Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Molecular Microbiology* **56**(3): 836-844.

Úbeda, C., Maiques, E., Tormo, M. Á., Campoy, S., Lasa, Í., Barbé, J., Novick, R. P. & Penadés, J. R. (2007b) SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol Microbiol* **65**(1): 41-50.

Úbeda, C., Tormo, M. Á., Cucarella, C., Trotonda, P., Foster, T. J., Lasa, Í. & Penadés, J. R. (2003) Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol Microbiol* **49**(1): 193-210.

van Belkum, A., Melles, D. C., Peeters, J. K., van Leeuwen, W. B., van Duijkeren, E., Huijsdens, X. W., Spalburg, E., de Neeling, A. J., Verbrugh, H. A. on behalf of the Dutch Working Party on Surveillance and Research of MRSA (SOM) (2008) Methicillin-Resistant and -Susceptible *Staphylococcus aureus* Sequence Type 398 in Pigs and Humans. *Emerg Infect Dis* **14**(3): 479-483.

Van Den Broek, I. V. F., Van Cleef, B. A. G. L., Haenen, A., Broens, E. M., Van Der Wolf, P. J., Van Den Broek, M. J. M., Huijsdens, X. W., Kluytmans, J. A. J. W., Van De Giessen, A. W. & Tiemersma, E. W. (2009) Methicillin-resistant *Staphylococcus aureus* in people living and working in pig farms. *Epidemiology and Infection* **137**(5): 700-708.

Van den Eede, A., Martens, A., Lipinska, U., Struelens, M., Deplano, A., Denis, O., Haesebrouck, F., Gasthuys, F. & Hermans, K. (2009) High occurrence of methicillin-resistant *Staphylococcus aureus* ST398 in equine nasal samples. *Veterinary Microbiology* **133**(1-2): 138-144.

van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., Couloux, A., Loux, V., Dervyn, R., Bossy, R., Bolotin, A., Batto, J. M., Walunas, T., Gibrat, J. F., Bessiäres, P., Weissenbach, J., Ehrlich, S. D. & Maguin, E. (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *P Nat Acad Sci USA* **103**(24): 9274-9279.

van Hijum, S. A. F. T., Zomer, A. L., Kuipers, O. P. & Kok, J. (2005) Projector 2: contig mapping for efficient gap-closure of prokaryotic genome sequence assemblies. *Nucleic Acids Res* **web-server issue**: W560-566.

van Valen, L. (1973) A new evolutionary law. *Evol Theory* (1): 1-30.

van Wamel, W. J. B., Rooijakkers, S. H. M., Ruyken, M., van Kessel, K. P. M. & van Strijp, J. A. G. (2006) The Innate Immune Modulators Staphylococcal Complement Inhibitor and Chemotaxis Inhibitory Protein of *Staphylococcus aureus* Are Located on Beta Hemolysin-Converting Bacteriophages. *J Bacteriol* **188**(4): 1310-1315.

Vandenesch, F., Naimi, T., Enright, M. C., Lina, G., Nimmo, G. R., Heffernan, H., Liassine, N., Bes, M., Greenland, T., Reverdy, M.-E. & Etienne, J. (2003) Community-Acquired Methicillin-Resistant *Staphylococcus aureus* Carrying Panton-Valentine Leukocidin Genes: Worldwide Emergence. *Emerg Infect Dis* **9**(8): 978-984.

Vath, G. M., Earhart, C. A., Rago, J. V., Kim, M. H., Bohach, G. A., Schlievert, P. M. & Ohlendorf, D. H. (1997) The structure of the superantigen exfoliative toxin A suggests a novel regulation as a serine protease. *Biochemistry* **36**: 1559-1566.

Vos, M. (2009) Why do bacteria engage in homologous recombination? *Trends Microbiol* **17**(6): 226-232.

Vybiral, D., Takác, M., Loessner, M., Witte, A., von Ahsen, U. & Bläsi, U. (2003) Complete nucleotide sequence and molecular characterization of two lytic *Staphylococcus aureus* phages: 44AHJD and P68. *FEMS Microbiol Lett.* **219**(2): 275-283.

Waldron, D. E. & Lindsay, J. A. (2006) Sau1: a Novel Lineage-Specific Type I Restriction-Modification System That Blocks Horizontal Gene Transfer into *Staphylococcus aureus* and between *Staphylococcus aureus* Isolates of Different Lineages. *J Bacteriol* **188**(15): 5578-5585.

Walther, B., Monecke, S., Ruscher, C., Friedrich, A. W., Ehricht, R., Slickers, P., Soba, A., Wleklinski, C.-G., Wieler, L. H. & Lubke-Becker, A. (2009) Comparative Molecular Analysis Substantiates Zoonotic Potential of Equine Methicillin-Resistant *Staphylococcus aureus*. *J Clin Microbiol* **47**(3): 704-710.

Wann, E. R., Gurusiddappa, S. & Hook, M. (2000) The fibronectin-binding MSCRAMM FnbpA of *Staphylococcus aureus* is a bifunctional protein that also binds fibrinogen. *J Biol Chem* **275**: 13868-13871.

Warr, G. W., Magor, K. E. & Higgins, D. A. (1995) IgY: clues to the origins of modern antibodies *Immunol Today* **16**(8): 392-398.

Watkins, G. H. & Jones, J. E. T. (2007) Mastitis and contagious agalactia. In: *Diseases of Sheep, Fourth Edition.* I. D. Aitken (ed). Oxford, UK: Blackwell Publishing Ltd.

Weigel, L., Clewell, D., Gill, S., Clark, N., McDougal, L., Flannagan, S., Kolonay, J., Shetty, J., Killgore, G. & Tenover, F. (2003) Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. *Science* **302**(5650): 1569-1571.

Wilder-Smith, A. & Memish, Z. (2003) Meningococcal disease and travel. *Int J Antimicrob Ag* **21**(2): 102-106.

Wilgenbusch, J. C. & Swofford, D. (2003) Inferring Evolutionary Trees with PAUP. In: *Current Protocols in Bioinformatics.* A. D. Baxevanis, L. D. Stein, G. D. Stormo & J. R. Yates III (eds). New York: John Wiley & Sons.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C., Ochman, H. & Achtman, M. (2006) Sex and virulence in *Escherichia col*i: an evolutionary perspective. *Mol Microbiol* **60**(5): 1136-1151.

Wirth, T., Wang, X., Linz, B., Novick, R. P., Lum, J. K., Blaser, M., Morelli, G., Falush, D. & Achtman, M. (2004) Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: Lessons from Ladakh. *P Nat Acad Sci USA* **101**(14): 4746-4751.

Woodford, N. & Livermore, D. M. (2009) Infections caused by Gram-positive bacteria: a review of the global challenge. *J Infection* **59**(S1): S4-S16.

Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* **32**(4): 569-577.

Yamaguchi, T., Hayashi, T., Takami, H., Nakasone, K., Ohnishi, M., Nakayama, K., Yamada, S., Komatsuzawa, H. & Sugai, M. (2000) Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*. *Mol Microbiol* **38**(4): 694-705.

Yamaguchi, T., Hayashi, T., Takami, H., Ohnishi, M., Murata, T., Nakayama, K., Asakawa, K., Ohara, M., Komatsuzawa, H. & Sugai, M. (2001) Complete Nucleotide Sequence of a *Staphylococcus aureus* Exfoliative Toxin B Plasmid and Identification of a Novel ADP-Ribosyltransferase, EDIN-C. *Infect Immun* **69**(12): 7760-7771.

Yanagi, H. & Sheretz, R. J. (2009) Infections of Intravascular Catheters and Vascular Devices. In: *Staphylococci in Human Disease, 2nd edition.* K. B. Crossley, K. K. Jefferson, G. L. Archer & V. G. Fowler (eds). Oxford, UK: Wiley-Blackwell.

Yang, L., Zhan, L., Han, H., Gao, H., Guo, Z., Qin, C., Yang, R., Liu, X. & Zhou, D. (2010) The low-salt stimulon in *Vibrio parahaemolyticus*. *Int J Food Microbiol* **137**(1): 49-54.

Yang, Z. (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555-556.

Yang, Z. & Nielsen, R. (2002) Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol Biol Evol* **19**(6): 908-917.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. (2000) Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics* **155**(1): 431-449.

Yang, Z., Wong, W. S. W. & Nielsen, R. (2005) Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol Biol Evol* **22**(4): 1107-1118.

Yarwood, J. M., McCormick, J. K., Paustian, M. L., Orwin, P. M., Kapur, V. & Schlievert, P. M. (2002) Characterization and Expression Analysis of *Staphylococcus aureus* Pathogenicity Island 3. Implications for the Evolution of Staphylococcal Pathogenicity Islands. *J Biol Chem* **277**(15): 13138-13147.

Yarwood, J. M. & Schlievert, P. M. (2003) Quorum sensing in Staphylococcus infections. *J Clin Invest* **112**(11): 1620-1625.

Yilmaz, A., Shen, S., Adelson, D. L., Xavier, S. & Zhu, J. J. (2005) Identification and sequence analysis of chicken Toll-like receptors. *Immunogenetics* **56**: 743-753.

Zadoks, R., van Leeuwen, W., Barkema, H., Sampimon, O., Verbrugh, H., Schukken, Y. H. & van Belkum, A. (2000) Application of pulsed-field gel electrophoresis and binary typing as tools in veterinary clinical microbiology and molecular epidemiologic analysis of bovine and human *Staphylococcus aureus* isolates. *J Clin Microbiol* **38**(5): 1931-1939.

Zadoks, R. N., van Leeuwen, W. B., Kreft, D., Fox, L. K., Barkema, H. W., Schukken, Y. H. & van Belkum, A. (2002) Comparison of *Staphylococcus aureu*s Isolates from Bovine and Human Skin, Milking Equipment, and Bovine Milk by Phage Typing, Pulsed-Field Gel Electrophoresis, and Binary Typing. *J Clin Microbiol* **40**(11): 3894-3902.

Zdobnov, E. M. & Apweiler, R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**(9): 847-848.

Zerbino, D. R. & Birney, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821-829.

Zetola, N., Francis, J. S., Nuermberger, E. L. & Bishai, W. R. (2005) Community-acquired methicillin-resistant *Staphylococcus aureus*: an emerging threat. *The Lancet Infect Dis* **5**(5): 275-286.

Zhou, J., Enright, M. C. & Spratt, B. G. (2000) Identification of the Major Spanish Clones of Penicillin-Resistant Pneumococci via the Internet Using Multilocus Sequence Typing. *J Clin Microbiol* **38**(3): 977-986.

Appendices: published papers